2.6. *Development of techniques.* Experiments (or tests or tutorial sessions) of the type described in this chapter are laborious but marginally possible. If the topic of the CET conversation is extended to cover a few days worth of continual study or the equivalent content of a two month course the arrangement is completely unmanagable. Moreover, the English language dialogue is, in many ways, embarrassing; it lays the results open to criticism from anyone who has not seen the experiment in progress and the personality biases that do exist are prone, in the long run, to hinder rather than help the student. It was thus decided to employ a language, with rather greater richness than the stilted form of English actually employed by the participant experimenter, which is mechanised as part of an experimental facility.

The facility is called CASTE (Course Assembly System and Tutorial Environment). It is described, after some orientating and recapitulating preliminaries, in Chapter 4. The participant experimenter is usually replaced, in this facility, by a CET heuristic or a tutorial heuristic (Chapter 1, Section 1.5) or, rather, the CET heuristic or tutorial heuristic is executed by a processor consisting of a general purpose computer and a collection of special purpose equipment. The heuristic is one participant, and the subject is another participant. The knowables are represented to both participants by a dynamic and continually visible display. The CET heuristic underlies any tutorial heuristic (as in Chapter 1, Section 1.5 the latter heuristic is a specialised form of the CET) and it will be shown that operation of the CET heuristic guarantees a series of occasions on which (in the framework of an experimental contract) there are understandings. Hence, the system is relativistic; the external observer examines Participant A relative to Participant B in the context of a conversational domain.

2.7. The mierotheory underlying a relativistic psychology consists in a theory of conversations (which is discussed in Chapter 5 and Chapter 6 together with a subtheory of participants) and a theory of knowledge and tasks that may be done (Chapter 7, Chapter 8 and Chapter 9).

Chapter 4. A Realistically Sized Conversational System

To get the discussion under way I shall make some intuitively plausible statements which, on closer inspection, are of a very dubious nature. None of them has an exact meaning at the moment, though all of them are eventually made precise. Their immediate but superficial reasonableness is chiefly due to the fact that they bear upon facets of the previous discussion.

First, it is maintained that any psychological experiment is a conversation between two or more participants A and B on a series of topics that form a conversational domain. Most of the argument deals with conversations between two participants. Only one participant (A) is a respondent (the word "subject" is avoided henceforward, since it is also necessary to use "subject matter"); the other, (B), is a participant experimenter or agent. For example, A might be an interviewee: or an examinee: B an interviewer or an examiner. Or A might be a student and B a teacher; if so the conversation is a tutorial. This is an important case in its own right. But, whether or not A and B have the official status of student and teacher, it is invariably true that they learn as a result of conversing about the domain (this is a condition; the domain contains something to be discovered or learned about, whatever roles such as respondent/interviewer or student/teacher are assigned to A and B). Although A is always human and B may be human, there are interesting cases, which are stressed because of their clarity, in which B is executed by a machine. In either case the A, B conversation takes place across an interface which separates A and B, which serves as their communication medium and at which position an external observer (in contrast to a participating experimenter in the role of B) may scrutinise and record A, B transactions.

The conversation takes place in a potentially formalisable language L. Although L can be a spoken or written language, in the special and clearcut cases where B's role is executed by a machine L is a mechanical language. If so L expressions typically consist in sequences of graphic displays, signalling events, or responses such as building a model, solving a problem, or writing a computer program. Whether spoken, written or mechanised, L must be rich enough to accommodate fairly sophisticated transactions: in particular, L is a command and question language not just an assertoric language.

For reasons that will be exhibited later, L is also stratified into at least two levels of discourse: $L = L^1, L^0$. Until further notice, it is useful to identify the levels of discourse L^1, L^0 with the levels of control (Lev 1, Lev 0) which featured in the earlier discussion. This connotation, though incomplete, is never positively misleading.

It follows that commands and questions may be issued either at L^0 or at L^1. At level L^0 commands are of the form "Do something" or "Solve a problem"; questions are of the form "Give an explanation". In contrast, at level L^1, commands are of the type "Learn to solve a problem (i.e. construct a process that solves it)" and questions call for an explanation of how the process was constructed (or, sometimes, reconstructed). The choice of how or why questions as the basic interrogations either at L^0 or L^1 is deliberate and will be justified in context; "which" and "what" questions, usually regarded as more elementary and tackled first, are readily obtained as special cases of the how and why question forms.

The A, B conversation is set in a normative framework. It takes place because A and B engage in an experimental contract (sometimes, more aptly, a tutorial contract) according to which they agreed to abide by the vocabulary and syntax of L: to interpret L with respect to the conversational domain (L semantics) and to entertain rules or classes of intentions (L pragmatics). In other words, the integrity of the L conversation depends upon rule obedience and the fact that certain rule obediences interlock. As the discussion proceeds it will become increasingly evident that the same comment applies to the integrity of A and B as well; in fact, A and B are classes of intentions or, in the terminology of computation rather than logic, they are unitary entities insofar as they are characterised by repertoires of procedures (where a "procedure" means either a serial program, a nondeterministic program, or a "fuzzy" algorithm). Because of this it is often important to distinguish between the participants A, B and the processors that execute these repertoires of procedures. The processors are designated α and β; moreover, the interface between α and β is distinguished as a further processor *i*. For unlike a standard behavioural experiment where the organism/environment boundary is set in terms of processors that are assumed to be in one to one correspondence with procedures proper to each one (α, A) and (β, B) the boundary between the participants A, B in a conversation can usually be drawn only if A is *not* uniquely correlated with α, B with β. We shall gloss the distinction between procedure class and processor, until lack of it becomes a nuisance, by referring to human beings (meaning, ambiguously, either human brains or the mindfully human procedures they execute) and by referring to systems (meaning, in general, either computing machines or the procedures they execute). Further, until the contrary is stated, the labels A and B will be used sloppily, with reference to human beings, or to systems. But the caveat should be kept in mind; really, A is a procedure class under execution (i.e. a role); B is a procedure class under execution (i.e. another role or, especially if it is executed by a machine, a heuristic).

Under these circumstances, it is possible to express the teachback condition (described earlier) as a type of complex agreement, henceforward called an understanding, which is approximated by the picture in Fig. 1. $\underline{Proc_A^0}\,i$ represents a procedure in A's repertoire that explains a topic labelled i; $\underline{Proc_A^1}\,i$ is a procedure in A's repertoire that explains how $\underline{Proc_A^0}\,i$ is learned and which, if applied to $\underline{Proc_A^0}\,i$ itself, may be regarded as A's re-explanation or A's reconstruction, or A's memory of topic i; $\underline{Proc_B^0}\,i$ is a procedure in B's repertoire also for explaining topic i and $\underline{Proc_B^1}\,i$ is a procedure in B's repertoire for explaining how $\underline{Proc_B^0}\,i$ is learned.

Suppose that A and B jointly address their attention to Topic i, then their agreement that they explain topic i in the same way is represented by the continuous, level L^0, loop in Fig. 1 and will eventually be formalised (when officially tenable meanings have been given to these terms) by expressions like:

$$\underline{Proc_A^0}\,i\,(X) \iff \underline{Proc_B^0}\,i\,(X)$$

where X is a domain on which a $\underline{Proc}^0$ is able to operate and "$\iff$" stands for "isomorphism" or one to one correspondence. In words "A's model for an explanation of topic i is isomorphic (has the same form as) B's model or explanation of topic i".

Similarly, the upper or L^1 loop in Fig. 1 signifies the fact that

$$\underline{Proc_A^1}\,i\,(\underline{Proc_A^0}\,i) \iff \underline{Proc_B^1}\,i\,(\underline{Proc_B^0}\,i)$$

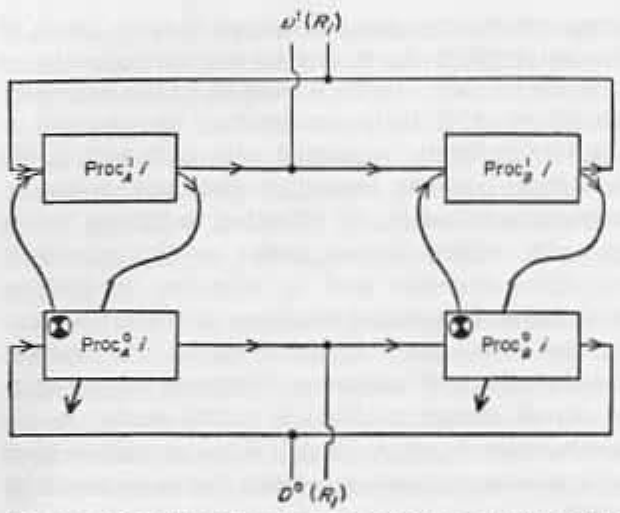


Fig. 4.1. Specific conversational transaction of explanation and "Teachback". Explicitly identify "⌛" with "description of" and "♀" with "reproduction of construction of". On given occasion, student/teacher point to one topic relation R_i in subject matter, R. $Proc_A^1 i$, $Proc_B^1 i$, $Proc_A^0 i$, $Proc_B^0 i$, are concepts (procedures) for R_i that will belong to repertoires; π_A^1, π_B^1, π_A^0, π_B^0, student builds $Proc_A^0 i$ (which is given in $D^0(R_i)$); perhaps with help of teacher using $Proc_B^0 i$; this concept may be used to satisfy or bring about R_i if R is instantiated. Student explains R by describing (at higher level) $Proc_A^0 i$; in the "Teachback" operation he also describes/explains how he brought $Proc_A^0 i$ into existence (permissible types of construction are given in $D^1(R_i)$). This explanation of how $Proc_A^0 i$ is built is the execution of $Proc_A^1 i$. If it is internalized, cycle of reproduction uses $Proc_A^1 i$ to reconstruct $Proc_A^0 i$. Teacher gives assistance, using $Proc_B^1 i$.

or, in words, that "A's method of learning (i.e. internally modelling or reconstructing $Proc_A^0 i$ and of explaining how he does so) is isomorphic to B's method of learning $Proc_B^0 i$ (i.e. internally modelling or reconstructing $Proc_B^0 i$) and of explaining how he does so.

If A and B do, on some occasion, n = 0, 1, . . .n, n + 1, ...N jointly address their attention to topic i and if the double layered agreement symbolised by Fig. 1 does hold, then we refer to the resulting stable configuration as a condition of "A, B, understanding a topic i expressed in L" or just of "understanding". One of the main tenets of this theory (amply confirmed by the empirical data) is that understandings can be detected in a certain type of conversation and used (amongst other things) to

demarcate occasions. The bulk of this chapter contains a very factual discussion of an experimental facility in which understandings are systematically detected and their development investigated.

At occasion n, we refer to $Proc_A^0 i$ as A's concept of topic i. (It is essential to our theory, as it is to the sensical interpretation of "concept acquisition" or "concept learning" as these phrases are generally employed, that a concept should be a procedure for doing something, not a class, that is somehow lodged in a mental register.) Further, if topic i is understood, then we refer to $Proc_A^1$ as A's memory of topic i or A's concept of A's concept of topic i (A's reconstruction and replication of $Proc_A^0$ i).

In view of the previous discussion it seems reasonable to contend that the entities people know and learn about are relations; similarly, that whatever people do is intended to bring about or satisfy relations; hence, that a topic i is a relation R_i. It turns out, on deeper examination, that if this point of view is taken wholeheartedly, then it commits us to a serious reappraisal of several fundamental choices (on the part of scientists) concerning the way they opt to look at their world. This matter is broached again in the next chapter; immediately, the relational identification seems innocent enough. For example, functional psychologists make no fuss at all about the fact that in perceptual-motor skills a display-control relation is learned and that this relation characterises the task in hand. As hinted earlier in this paragraph, our theory maintains that the basic knowables and do-ables are relations (in this respect, the functional psychologists are right) but it also maintains that this view is of some consequence and liable to disrupt the status quo in other areas. At all events, a relationally oriented point of view will be adopted, henceforward, to await justification later. By extrapolation, any coherent area of topics, the conversational domain in particular, is a related collection of relations; it is denoted R and a topic relation is denoted R_i in R.

The procedural argument can be generalised in line with these sentiments. Generally, $Proc_A^0 i$ is a process which produces or brings about the i-th relation in an appropriate domain or, if this relation holds at a given instant, which reproduces it so that it continues to hold later. By the same token, $Proc_A^1 i$ is a process that builds $Proc_A^0 i$ by writing the procedure $Proc_A^0 i$ into any uncommitted storage locations of a processor, or if $Proc_A^0 i$ already exists, then a process that reproduces $\underline{\text{Proc}}_A^0$i, as a replicate procedure. These processes are mutually dependent, though the latter may be highlighted as "A's learning". But learning, if it is externalised, becomes "teaching" and the entwined pair of processes in A and B is "teachback".

Like any other entity that is operated upon by a procedure, a topic relation, R_i, must be described; to or by or for the procedure and in whatever language is employed to represent the procedure; here, in L. Since L is stratified by edict ($L = L^1, L^0$) there are at least two descriptions of any conversational domain, R; designated $D^1(R)$ and $D^0(R)$. Of the two, $D^0(R)$ is a permission giving structure that states how the R_i in R may be satisfied, or modelled or explained, or brought about. Used descriptively, $D^0(R)$ says what may be done; used prescriptively, to generate L^0 commands and/or L^0 questions, $D^0(R)$ is a basis for saying what must be done, by way of constructing models or of answering questions. $D^1(R)$, on the other hand, is a permission giving structure indicating the generally very many ways in which R_i may be synthesised from other topic relations in R. Used descriptively, $D^1(R)$, states what may be known or learned; used prescriptively, it is a basis for saying what must be learned; that is, what $\underline{\text{Proc}}^0$s should be constructed and how they should be built by $\underline{\text{Proc}}^1$s from other $\underline{\text{Proc}}^0$s.

Just as the repertoire of procedures which make up a participant A or B is executed and embodied in some processor α or β (we have reserved the right not to be dogmatic over which processor) so, also, any description $D(R) = \langle D^1(R), D^0(R) \rangle$ that is accessible to the participants must be inscribed and embodied in something; generally, in the interface *i*. The configuration is approximated by Fig. 2 in which π_A^1 denotes A's repertoire of memories, $\underline{\text{Proc}}_A^1$; π_A^0 denotes A's repertoire of concepts, $\underline{\text{Proc}}_A^0$; π_B^1 denotes B's repertoire of $\underline{\text{Proc}}_B^1$ and π_B^0 denotes B's repertoire of $\underline{\text{Proc}}_B^0$. The description $D(R) = \langle D^1(R), D^0(R) \rangle$ is accessible (perhaps conditionally) to A and B. The physical inscriptions at the interface are called the entailment structure ES(R) corresponding to $D^1(R)$ and the task structure TS(R) corresponding to $D^0(R)$. In general, the phrases "entailment structure" and "task structure" are used sloppily to designate either $D^1(R)$ or ES(R); either $D^0(R)$ or TS(R).

Suppose, as previously suggested, that A and B do not both know all of the conversational domain (at least one of them does

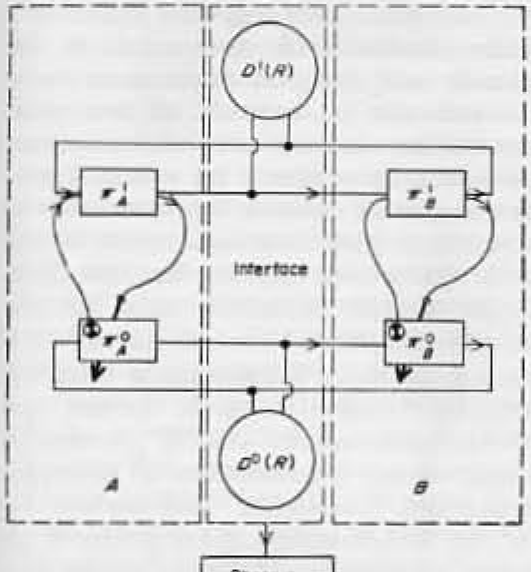


Fig. 4.2. Students = A, Teacher = B; levels of discourse indexed 1,0. $D^0(R)$ = Lower level description of subject matter relation. $D^1(R)$ = Higher level description; Student's and Teacher's higher and lower level repertoires of concepts (procedures) designated π_A^1, π_B^1, π_A^0, π_B^0, external observer has access to the conversational interface and observes dialogue.

not have $\underline{\text{Proc}}_A^0$i or $\underline{\text{Proc}}_B^0$i, for all R_i in R). Suppose, also, that there is a cooperative condition such as "neither A nor B alone (using their own Procs) can learn the whole of R but both of them may do so if they act mutualistically". For most purposes a much weaker form of cooperation is enough "that A finds it hard to learn all of the topic relations in the domain without the assistance of B" (the reverse is also required for a symmetrical conversation in which the participants play the same roles but not for the asymmetric case typified by a tutorial). Next, suppose that A's and B's repertoires of procedures are executed in some processor(s). Finally, suppose that the execution is synchronised in respect of a particular topic relation such as R_i, so that A and B address their attention to R_i on occasion n. If so, then Fig. 2 reduces, momentarily, to Fig. 1 and it is possible to detect an A, B understanding that ends occasion n. At that juncture we may insist that A and B shall attend to and learn about some other part of the conversational domain. The "other part" may be a topic of A's (B's) own selection provided (a) that A (B) does not know the chosen topic relation already and (b) that (in a sense to be discussed) A and B are both able to learn it. At this point $n \rightarrow n + 1$; and so the conversation continues over a finite number of occasions until all the potential knowables in the conversational domain are exhausted. As a temporary measure, it is convenient to think of occasions, n, as similar to trials in an experiment. But, as the discussion proceeds, it will become obvious that this interpretation is far too simple minded. Occasions mark out the inception and completion of a complicated process; namely A, B synchronisation (or equisignificantly, A, B information transfer).

A piece of equipment, the "Course Assembly System and Tutorial Environment" (with the acronym "CASTE") is used to keep track of occasions and regulate a conversation of this type, for example, by mediating the "insistences" and testing the conditions (a) and (b) of the last paragraph. As a premium, an external observer can record all segments of the L dialogue that cross the CASTE interface, i.e. those that signify understanding. CASTE is also used to regulate more elaborate conversations (course assembly processes in which the conversational domain evolves).

1. The CASTE facility

CASTE is an essential tool for studying conversations and is a clear embodiment of many parts of the theory. Its operation is most readily described by reference to a specific series of experiments (reported in detail in Pask and Scott, 1972, 1973). The respondent in any of these experiments plays a student-like role and is usually called "the student" or A. For the experiments of particular concern, A interacts with a heuristic, B, which, depending upon the conditions, is either designed to externalise normally private aspects of cognition as stretches of L dialogue (i.e. an "unbiased" conversational heuristic) or it is a modified form of this heuristic, designed to encourage faster or more effective learning (recall that A learns something if any conversation is maintained). Other experiments, briefly noted because they add to data about strategies of learning, involved more rigid teaching operations.

The unbiased conversational heuristic is of special consequence

for several reasons. It is a realisation for complex subject matter, of the cooperative externalisation technique, (CET) previously used for much more clearly structured types of mental activity. In deference to this fact, it is called a CET heuristic. The CET heuristic furnishes the least restrictive situation in which understanding can be unequivocally detected and the occasions in a conversation demarcated. Finally, all other CASTE heuristics are based on the CET heuristic. For example, the tutorial conversation (the main alternative condition) is obtained by imposing macro-theoretic boundary constraints upon CET operation and by this means regulating A's uncertainty (which is determined, on line, by a series of confidence estimates).

For all of the experiments under discussion, the subject matter of the conversational domain, R, was elementary probability theory. The material covers, at appreciable depth, the topics usually clustered beneath "Bayesian Experiments" (generally one trial experiments, but "conditional probability" is included). Though several comparably large subject matter areas have been programmed into CASTE (as a description $D(R) = \langle D^1(R), D^0(R) \rangle$ and the tutorial data required to back it up) probability theory is the most fully studied and it is also one of the most interesting. Since probability theory is an applied science, its exercise calls for continual cross reference between a real world (of experimental designs, of frequency counts) and an abstract world containing mathematical constructs (structural models of experimental observations and measure theoretic models stipulating the "probability numbers" that may be assigned to models and the abstract inferences possible in these terms). But a student who comes to grips with probability (not just "frequencies" or "probability numbers", in isolation) is able to manipulate each kind of reality and to make valid statistical inferences that place the tangible in correspondence with the abstract. All of our students did so, and retained the skills they had acquired very satisfactorily.

The knowable topic relations in a subject matter of this type (almost any applied science) are, a priori, more difficult to represent than the knowables of an homogeneous subject matter (symbolic logic for example, or mathematics, where the topic relations belong to one universe of discourse). Since an applied science (or strictly, a theory of applied science, encoded in $D^1(R)$) necessarily spans several universes of discourse its representation is replete with strict analogies in addition to some analogy relations that are less strict. The fact that probability theory can be represented in a nontrivial form and that the representation proves effective is interesting in its own right and, incidentally, makes it possible to quote specific examples, culled from the experiments, that exhibit very general and important notions.

1.1. *CASTE layout*. The *interface* is shown in Plate 1. It is a room in which the student is free to move from one working area to another. The large display depicts a map of the subject matter and, in operation, is marked by clusters of signal lamps to indicate various features of the preceding dialogue. This photograph was taken when CASTE was set up for the conversational domain "elementary probability theory" (used, throughout, as a main example), and Plate 3 shows in detail, the inscriptions on the large

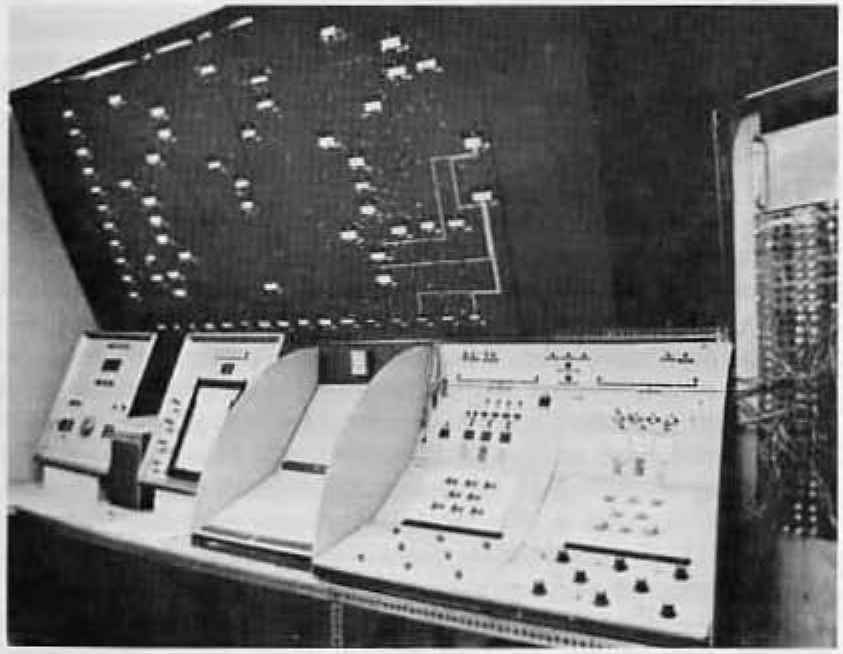


Plate 1. Student's working station in CASTE. The large display with clusters of marker lamps around each labelled topic is the entailment structure. By wall on left, tutorial data files, student access console. Next unit: Belief and Opinion Sampling System (BOSS) and files. Next unit: STATLAB (device for giving demonstrations and building models as non verbal explanations). Right hand unit: descriptor programming board and part of local processor.

display. Towards the lower right hand part of Plate 1, there is a modelling and simulation facility, STATLAB, also part of the interface, via which modelling operations are performed as the explanatory answers to questions about probability theory. The student gains control of the interface through an accessing console, shown in the lower left part of Plate 2. Between STATLAB and the accessing console is a confidence estimation device, BOSS, (Belief and Opinion Sampling System). A change in subject matter (for example, from probability theory to logic or mathematics or history) changes the fascia in front of the large display and its preprogrammed supporting circuitry. It also changes the data stored in various electronic and pictorial data bases and the modelling facility (which is specific to the subject). The accessing console and its connections are not modified.

The L description of the conversational domain (R) is stratified,

Plate 2. Experimenter's working station in CASTE. The teletypewriter is auxiliary device for accessing supervisory program executed in time shared system (local processors, various units in picture, are connected through direct transducers to Modem and time shared system). Similarly, the meters for reading confidence estimates and values of other variables are auxiliary devices.

$D(R) = \langle D^1(R), D^0(R) \rangle$ and the description is embodied in the interface $\langle ES(R), TS(R) \rangle$. Of these parts, ES(R) is the large display, its indexing and the electronics that back it up; TS(R) is the modelling facility STATLAB together with tutorial data files (some consulted manually and some through a random access projector). The interface is thus a dynamic system, consisting in

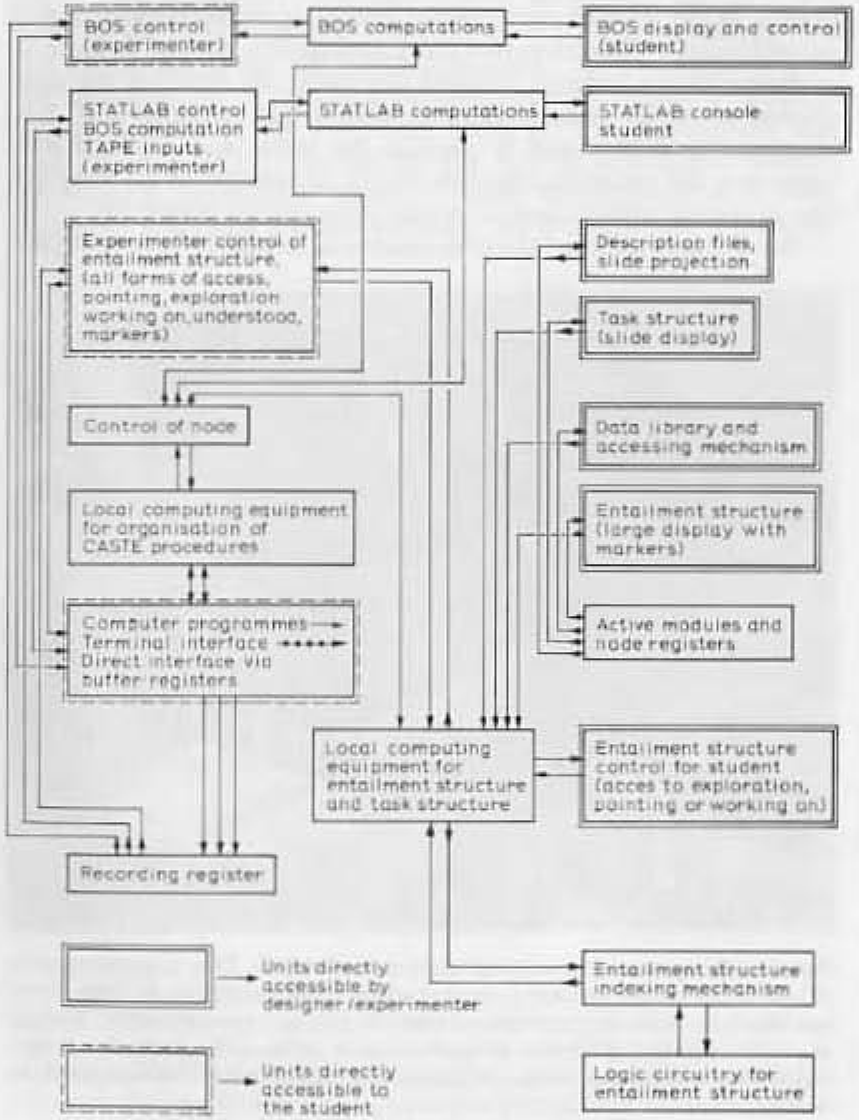


Fig. 4.3. Composition of CASTE, showing main units.

two machines with their own display and control arrangements that are constrained by $D^1(R)$ and $D^0(R)$ respectively.

Depending upon the experimental conditions, B is either a CET heuristic or a tutorial uncertainty regulation heuristic. The heuristic can be executed within processor β, but parts of it may also be executed by a human experimenter, who sits in the control room of CASTE (Plate 2) and observes the student through the viewing window. If the experimenter does execute part of the heuristic his L interaction with the student (A) is confined to the L transactions B permits by the role of B. His main role is to act as an external observer, and these roles must be carefully separated. To secure this distinction it is useful to image the external observer as a different human being sitting in the same control room, (i.e. "external observer") signifies an experimenter who "looks on" and does not engage in L conversation.

A portion of the β processor is a remote (standard, time shared) computer facility accessed through a modem and phone line, either from buffer registers in the interface or through the terminal (Plate 2). For many purposes it is convenient for the experimenter to interpolate statements through the terminal even

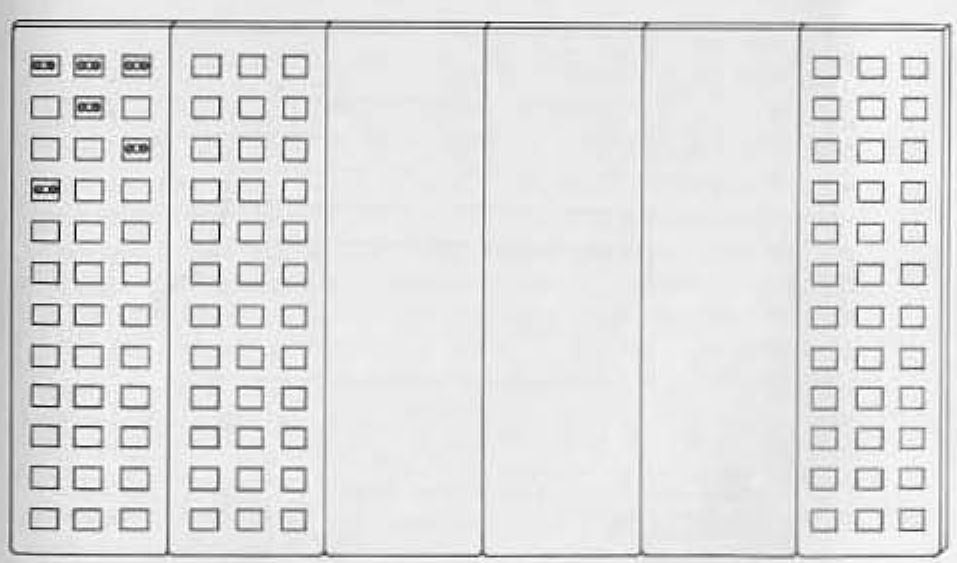


Fig. 4.4. Layout of frame for entailment structure. Each board contains 36 standard positions which may be covered by a custom built facia or a marker board.

though most of the B heuristic is mechanically executed; some of it locally and some of it by the time shared computer. Under these circumstances, for example, the experimenter can "off load" onerous checking and recording tasks whenever he is overtaxed; nearly all the updating storage needed to instrument "off loading" is available in the interface equipment.

The organisation of the system is shown in Fig. 3 and Fig. 4. The captions furnish all the technical details necessary for this discussion (operating routines are given as annotated programme listings in Appendix D and Appendix E; details of the processors and auxiliary equipment circuits are described in a technical document; the CASTE manual 1971).

1.2 *Experimental background.* The following comments give a fair idea of how respondents acting as students see the CASTE facility and are presented with its rules of operation.

All respondents were anxious to act as students and to learn probability theory, the subject matter shown in Plate 3. They were between 18 and 25 years old; mostly attending technical or art

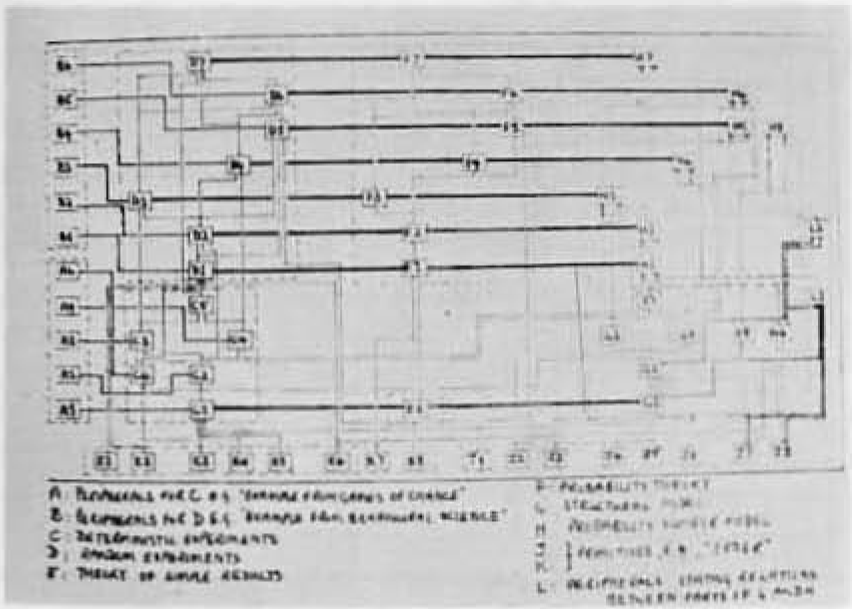


Plate 3. Entailment structure for thesis on elementary probability theory showing assignment of descriptor values.

schools; some garnered from cafe society in Richmond and its neighbourhood. The majority of respondents had a background in the humanities and were considering or embarking upon courses in social science, psychology, or economics that involved acquaintance with the nature of statistical techniques and/or the formal representation of processes, systems, etc. Prior to entering the experiment, potential respondents were interviewed and told of certain commitments, to which they would be held very rigidly; for example, attendance at the laboratory for lengthy periods during the experiment and attendance for examination later. The time spent was paid. The level of remuneration, though sufficient to justify the expenditure of useful time, (7/6d = 37½p per hour) was insufficient to be attractive in its own right (compared with a typical vacation job, the work would be hard and regular, though interesting). Certain filtering techniques were employed, during the interview, to obtain a heterogeneous sample of cognitive styles, but the main rejection criteria was above average numerical ability, as judged by a standard paper and pencil test and confirmed by discussion. It is fair to say that all of the respondents were of *below* average numerical aptitude, that some were virtually innumerate and that several had a positive aversion to mathematical ideas.

Accepted candidates came to the laboratory for pre-testing (relevant to the experiment but not to the CET heuristic as such). They were introduced at the same session to the equipment shown in Plate 1 and Plate 2; its operation was demonstrated and detailed. The only presentational difficulty was due to the mechanical impedimenta. Because of certain ingrained preconceptions about computing machines, it was hard to convince the students that no regulations are concealed or implied; that is, if B is a CET heuristic (or, in fact, any conversational heuristic) there is much less restriction than they would normally meet with in a classroom or even a sympathetic tutorial. The notion of ES(R) (Plate 1 and Plate 3) as a specification of what may be learned was explained, with especial emphasis on the fact that many ways of getting to know are possible and permissible. It was further determined that all of the students could understand the lowermost topic relations in ES(R).

This point requires a little clarification. Most of our respondents understood these and only these topics and this circumstance, whilst experimentally provident (since it allows us to sample as much learning as possible) does involve a couple of harmless tricks. First, the entailment structure was constructed to fit the known intellectual norms of the population, as determined by lengthy pilot experiments using different but overlapping subject matter. Next, the lowermost nodes were fitted to the cultural idiom, insofar as they correspond to concepts that generally exist but are not generally named; hence, they can be specified by definition. If not (and there were several exceptions) the topics in question were taught as laboriously as necessary to the respondent and he was required to teach them back to the experimenter. In either case, care was taken to ensure that all these topics could be explained and that their derivation could be explained. The explanations were verbally elicited but rigorously adjudicated with respect to an entailment structure considerably larger than the ES(R) shown in Plate 3. The method employed was identical with the teachback technique used in the conduct of earlier studies of learning strategies (Pask and Scott, 1972), and in Chapter 3.

1.3. *Salient features of the display.* The topic relations R_i are represented to the student as labelled nodes in the entailment structure. At the top of the main display (of ES(R) or D^1(R)) is a cluster of nodes called a Head (nodes D, 1; F, 1; H, 1 in Plate 3) since they stand for the analogous (centrally located) topics bound together by probabilistic inferences made between the factual world (left side) and the abstract world (right side). The lowermost nodes, as mentioned before, represent topic relations that any student who enters the system for this experiment is able to comprehend (this, incidentally, is not a general restriction). In between the head and the lowermost nodes there are nodes representing other topic relations and the student is told that he may gain access to any of them if B permits it; he definitely is not encouraged to work sedately upwards from the lowermost topic relations with which he is already familiar.

All students received procedural instructions tantamount to rules for using L. The written instructions look complicated. They are almost as incomprehensible as the written out rules for a parlour game. But in practice, with the equipment at hand, the instructions are no more difficult to follow than the game rules are, once the physical paraphenalia is handled. Amongst other things, a student is told how to ostend parts of the subject matter, how to ask questions, and how to submit explanations (STATLAB

models) in reply to the questions B asks of him (A). The confidence estimation facility; the "Belief and Opinion Sampling System" or BOSS, is also explicated. It is used as a measuring device, only, by the CET heuristic; in order to determine the value of "macro scale" variables, by the uncertainty regulation heuristics. The L transactions covered by the instructions are shown in Table 1, together with their meaning in terms of ordinary language.

1.4. *States of nodes in display.* When CASTE operates, the student receives a continually updated account of his progress towards learning, understanding and explaining the Head topic relations through indicators on the main display.

Each node in the display is associated with electronic registers that retain data about the current state of the node. Any node may be in an (exclusive) state represented as a conjunction of the following quantities (defined in Section 2.6; see also Table 1); namely, *explore, aim, goal, subgoal, tag* and *understand.* The variable understand is peculiar insofar as a positive value (node marked as being understood) is a unique state and thus precludes the possibility that this node is an aim, a goal, etc. The state of each node at any instant is signified by the illumination of 3 differently coloured signal lamps, either continuously or intermittently. Values of these state variables depend upon a history of L transactions, over the entire learning session. Some of them (explore, aim, goal and subgoal) have values set by A with the sanction of B. The values of tagaim and understand are set by B only. Understand is used to mark the node of any topic relation for which A's understanding is manifest in the manner equivalent to "teachback". Understand, once set, is indellible under normal conditions and is uniquely signified by a continuously illuminated green signal lamp. Hence, for the ES(R) of Plate 1 or Plate 3, the condition that R(Head) is satisfied is seen as the green illumination of signal lamps on the nodes for R(<D, 1>), R(<F, 1>) and R(<H, 1>). Prior to this assessment, the student must have been asked to explain each of these topic relations and (before or after doing so) to explain how he learned the concept in question.

1.5. *Experimental contract with the respondent.* From a respondent's point of view "being a student" (i.e. assuming the role of A) means that he will either obey a command to explain R (Head) by a permissible method (if he can do so, the experiment is

TABLE 4.1
Main CASTE Transactions

Code name	Logical function	Method employed	Restriction imposed	Permission given	Mark used
Explore	Ostension and enquiry into description	Dials node index	None	Description of relational network underlying node is available	Intermittent white lamp
Aim For	Student indicates most distant appreciated node as his current learning goal	Dials node index and qualifies by aim for. Retained unless deleted	Node must not be understood. If aim is marked working set must be specified	Allowed to construct working set of nodes entailed by aim for	White signal lamp
Goal	Student gives goal description	Dials node index and qualifies by goal	Node must not be understood and must be entailed by aim node	Allowed to point at subgoal in working on set	Dim modulated red signal lamp
Work on	Work-set	Retrieved by heuristic	Members of working set may not be deleted, only changed into understood		Intermittent red signal lamp
Subgoal	Student gives subgoal statement	Dials node index and qualifies by subgoal	Node must be in working set but concurrent selection permitted	Given access to tutorial data but committed to giving explanation of topic relation of each node	Red signal lamp
TagAim	Marker used by heuristic in specification of legal working set				Special lamp
Understood	Marker that indicates concept and memory of node i	Determined by heuristic on reception of explanation in the appropriate form	Prohibit aim for or work on. Remove from working set	Provides image of topic relation understood and allows specification of legal working set	Fixed green signal lamp

completed since he knows "probability theory" at the outset) or, lacking the requisite knowledge to do so, that he will reinterpret the directive as a command to learn to obey R(Head) by some combination of methods permitted by ES(R). His learning and/or his modelling/explaining are governed by the rules and interpretation of L and certain special constraints imposed by the heuristic, B.

2. Overview of the system and its rationale

So far as a student is concerned the entailment structure is a map of knowledge in which he can move around by using the appropriate transactions. At any point it is possible, again by using appropriate transactions, to obtain data/demonstrations that delineate what may be done if a concept is known. However, in obtaining this data, the student becomes involved in various commitments to give explanations or make models and his performance is evaluated. The original map is marked by tokens showing both the student (A) and the heuristic (B) the path travelled; the topic relations currently regarded as understood, and so on.

Although this picture is adequate for the learner it must be enlarged before it offers much guidance to an observer who wants to know what is going on.

TABLE 4.2
Key for nodes of entailment structure (Plate 3)

The names of the descriptors (1st, 2nd, 3rd, and 4th) correspond to the notation in Chapter 4, where only 1st and 2nd are relevant to the discussion. In the detailed account given in later chapters, all descriptors are relevant but the order of 1st and 2nd must be inverted as the reserved numerically valued, "First Descriptor" ($\underline{\text{Pred}}_1^1$) is the "2nd descriptor" in the table, and "1st" (in the table) may, for this purpose, be regarded as the "2nd descriptor".

Node index	1st	2nd	3rd	4th		Node index	1st	2nd	3rd	4th	
1	C	11	*	ϕ	Simple results	29	M_1	7	*	ϕ	Probability numbers (P. Nos.) of simple events
2	C	10	b	ϕ	Composite results	30	H	6	*	ϕ	P. No. of composite events
3	C	10	a	ϕ	Exclusive results	31	H	5	*	ϕ	P. Nos. of exclusive events
4	C	9	b	ϕ	Inclusive results	32	H	4	*	ϕ	P. Nos. of inclusive events
5	C	9	a	ϕ	Deterministic experiments	33	H	3	a	ϕ	Conditional P. Nos. of simple events
6	C	8	*	ϕ	Random results	34	H	3	b	ϕ	Conditional P. Nos. of composite events
7	D	7	*	ϕ	Frequency of simple results	35	H	2	*	ϕ	Probability theory
8	D	6	*	ϕ	Frequency of composite results	36	H	1	*	ϕ	P. No. of complement of composite event
9	D	5	*	ϕ	Frequency of exclusive results	37	J	12	*	1	Model/Real World distinction
10	D	4	*	ϕ	Frequency of inclusive results	38	J	12	*	2	Inclusive/Exclusive distinction
11	D	3	*	ϕ	Conditional frequency of simple results	39	J	12	*	3	Arithmetic operations
12	D	2	*	ϕ	Conditional frequency of composite results	40	J	12	*	4	Universe
13	D	1	*	ϕ	Random experiments	41	J	12	*	5	Set and subset
14	E_1	11	*	ϕ	Theory of simple results	42	J	12	*	6	Complementation
15	F_1	7	*	ϕ	Probability of simple results	43	J	12	*	7	Intersection
16	F_1	6	*	ϕ	Probability of composite results	44	J	12	*	8	Union
17	F	5	*	ϕ	Probability of exclusive results	45	K	12	*	1	Long-run stability
18	F	4	*	ϕ	Probability of inclusive results	46	K	12	*	2	Counting
19	F	3	*	ϕ	Conditional probability of simple results	47	K	12	*	3	Order
20	F	2	*	ϕ	Conditional probability of composite results	48	K	12	*	4	Qualities
21	F	1	*	ϕ	Probability theory	49	K	12	*	5	"One at once" definition
22	G	11	*	ϕ	Event set	50	K	12	*	6	Fractional numbers
23	G	10	*	ϕ	Composite events	51	K	12	*	7	Definition of probability relation
24	G	9	a	ϕ	Exclusive events	52	K	12	*	8	Definition of given knowledge
25	G	9	b	ϕ	Inclusive events	53	L	9	*	ϕ	Relation between logic of structure and arithmetic of P. Nos.
26	G	9	c	ϕ	Structural model	54	L	6	*	ϕ	Relation between P. Nos. and conditional P. Nos.
27	G	9	d	ϕ	Probability number	55					
28	G	8	*	ϕ	Complement of composite event						

2.1. *Conversational domain.* The chief inaccuracies of the student's eye picture lie at the level of the conversational domain, R, and its description $D^1(R)$ embodied in ES(R). The student (A) can afford to see $D^1(R)$ or ES(R) as a map of knowledge simply because the "map" is imbued with special properties by B's operation. An observer can only make sense of the entire system if he has a much more cogent image of R and its description $D^1(R)$. To draw this image in outline (it is detailed in later chapters) we need to look back at the origins of R and $D^1(R)$, and to comment upon how a theory (for example, an elementary theory of probability) came to be expressed in this form. In addition, it is necessary to look behind the map like display and to specify the data structures, encoded in the CASTE interface, that back it up.

2.2. *The Entailment mesh.* The graph in Plate 3 is an inscription of a pruned cyclic entailment mesh which stipulates what may be known. The nodes in the graph represent topic relations; that is, they are place holders for these relations. Arcs in the graph are of two kinds "$O \longrightarrow O$" and "$O \Leftarrow O \Rightarrow O$" of which the latter is a special condensed notation used to replace an otherwise hard to perceive plexus of "$O \longrightarrow O$" connections that represent an analogy. Each arc stands for a permitted relation between topic relations; a permitted method of getting to know (or of learning). A very large number of methods are permitted.

2.3 *Names of topics.* Any node has a label showing the name (in L^1) of the topic relation it represents. During construction a consistency condition is demanded, ensuring that names are in one to one correspondence with topic relations. The names pertinent to Plate 3 are listed in Table 2 where they are indexed according to the alphanumeric coordinates used to partition the nodes.

2.4. *Indexing.* The order of writing in Table 2 is given by a numerically valued index i of topic relations R_i. For this example (simple probability theory) i ranges from 1 to 56.

For any conversation on a fixed domain i ranges over a fixed set of topics and this arrangement is usually adopted in wholly tutorial applications of CASTE. In principle, however, i is the index of an open ended list to which topics are added from time to time.

2.5 *Coordinates of the display.* The alphanumeric coordinates represent values of *descriptors* or L^1 predicates. They are chosen

so that subsets defined by conjunction (for example <1st = G, 2nd = 5>) are sets containing at most one place-holder for each topic relation indexed by one value of i. Moreover, all place-holders and thus all topic relations R_i are accommodated in some unit set. Hence, it is always possible to uniquely ostend a relation by an appropriate L^1 conjunction, as you did when referring to Table 2 and as a student does via a dialling operation on his accessing console. But it is also possible to ostend loci that contain no place holder at the moment, for example, "<1st = E, 2nd = 5>" and the existence of this possibility is an essential part of the scheme.

2.5.1. The descriptors are the coordinates of a terrain; like the grids and contour lines of a map. As such, their values furnish little or no information about the ostended topic relation. They describe particular topic relations (as "*Value 1st, Value 2nd*") and it is entirely possible to use them for describing the structure of the terrain; for example, to say that one topic is "higher valued on 2nd than another" or to specify classes of topic relations as part of a plan, an assertion, or an hypothesis.

2.5.2. One descriptor (2nd in this case) has a chiefly geographical meaning; the value "1" of 2nd is higher in the display than a value "2" of 2nd. The student is told that this number assignment does not imply a definite ordering which he has to obey (he can, for instance, aim to deal with any topic as the first in a study sequence). The graph is not an hierarchical "knowledge structure", even though it has a quasi hierarchical appearance. For all that, the descriptor "2nd" does have a deeper connotation which will become clear on scrutiny of the construction process; its values reflect the way in which whoever designed the pruned entailment mesh chose to extract it from a more fundamental structure.

2.5.3. The descriptor, "1st", has a common language meaning and above its geographical meaning. Thus Gs are topic relations to do with abstract logical models; Hs are to do with abstract measures on these models; Cs are to do with real experimental situations; Ds are to do with the frequencies of experimental results and their ratios; Es are inference relations (strict analogies) between C topics or vice versa and Fs are inference relations (strict analogies) between D topics and H topics or vice versa. It is still true, of course, that a value of "1st" on its own conveys little or no information about what the topic relations are; but (like a value of "2nd" it does convey information about the structure of the topics and their organisation).

2.5.4. Other L^1 predicates may be (and, in fact, are) added to the description language. They may be logical functions of expressions in the original predicates, for example: "real/not" corresponding to "All Ds and all Cs/rest" or "Abstract/not" corresponding to "All Gs and all Hs/Rest". On the other hand they may be generated de novo (for example, the descriptor, "machine like" or the descriptor "conditional probability and information measure" or even "due to Bayes" and "invented prior to 1880"). For that matter, "1st", itself, is encoded from, and may be decoded into more elementary properties that are L^1 predicates (descriptors). In particular two further descriptors 3rd and 4th are used in Plate 3 and Table 2 to discriminate the lowermost nodes and certain peripheral nodes that are shown at the left of Plate 3 the values of A, B of 2nd.

Like 1st these descriptors are generally many valued but unary (1 place) predicates of an object variable i ranging over the nodes or else (like 2nd) such predicates ordered by integer. Clearly, an economic description of a path or a pattern on the graph involves many-place L^1 predicates or their equivalent; these are not usually formalised since the graph itself is displayed as a (many-place) descriptive statement in L^1.

Under certain circumstances, to be discussed, a student may generate new descriptors of his own choice during the learning process; also, he may use them and the existing descriptors to predicate values of i that have no referent in the initial set of 56 nodes.

2.5.5. L^1 descriptors describe the form of what may be known (not what may be known). Given this image, a student is able to explore a subject matter and plan to learn it, so that a topic can be examined in relation to others even though the student is ignorant of its content and even though he may not be in a position to start learning about the topic on the occasion when it is explored. Moreover, using the descriptors, a student can point cogently at whatever he currently aims to learn about, whether or not he has worked on topics in this neighbourhood.

There are an indefinite number of descriptors relevant to the form of any subject matter. But unless some of them are cited, learning, in the present sense, would be impossible and learning in any reasonable sense would be unobservable.

2.6. *Marker predicates.* The node or place holder for topic relation R_i is associated with tokens which can be deposited to mark its current condition. In CASTE, these tokens consist in a cluster of differently coloured signal lamps (Section 1.5) which may be fully or intermittently illuminated to indicate values of eight unary and two valued $\{1, 0\}$ L^1 marker predicates namely,

(1) Understood.
(2) Member of work set.
(3) Goal (candidates for work set).
(4) Subgoal.
(5) Aim for.
(6) TagAim.
(7) Explore.
(8) General transaction (a "spare" value, for the present discussion).

Restrictions are imposed by the L syntax which excludes certain combinations of these values; but the restrictions are peculiar to a heuristic. In any case, each node i is assigned a marker configuration as a result of the ongoing activity and it is always characterised by a conjunctive L^1 statement involving all of the marker predicates, its state.

Hence, the display is dynamic. The current pattern of marker configurations is a data store (Fig. 3, Fig. 4) from which A can check on his present status. Outputs, from the registers attached to these signal lamps (Plate 3, Fig. 3 and Fig. 4) inform the B heuristic of the current condition and B, in turn, renders certain data accessible or inaccessible to A as well as permitting and prohibiting certain moves on A's part (Fig. 3, Fig. 4).

(a) If and only if node i is explored can A receive values of the additional descriptors on R_i (i.e. those apart from 1st and 2nd).

(b) Only if node i is marked aim for can A receive information from an operator data base about how it is permissible to construct one relation from others he may know.

(c) Any node (or collection of nodes) beneath aim may be marked goal(s); meaning these are candidates for inclusion in workset (the B heuristic governs their examination as valid candidates possibly setting up TagAim statements in the process).

(d) If and only if a node is in workset, it can be addressed as a subgoal (if there are many goals, there may be many subgoals).

(e) If and only if node i is indicated as subgoal can A receive tutorial information about R_i (information about the relation itself, rather than the structure in which it is embedded).

(f) On receipt of this information A is bound to essay an explanation of R_i (a model that brings about R_i) as a result of which R_i may be marked as understood.

2.7 *Occasions and instants.* An occasion, indexed n, lies between changes of understanding which are heralded by a change in the marker predicate, understood.

2.7.1. When the conversational domain (hence the pruned cyclic entailment mesh) is fixed, as it is assumed to be, at the moment, an occasion is ended by a change in value of marker predicates from workset to understood. Any such transformations change the occasion number; $n \rightarrow n + 1$.

2.7.2. It is possible, even in a conversation on a fixed domain, to ostend notes that do not have a real valued index i (for example <E,5>). Such nodes are open to scrutiny; they exist in the display and electronically speaking in the interface equipment. Moreover, they can be described. But, as yet, they have no name or index. Fixed domain heuristics do not allow A to aim for, work on, or understand a node until it has been given a name; and, if a described but un-named node is ostended, the student is told that it is empty. The act of giving a number to such a node changes the conversational domain.

2.7.3. The systematic growth heuristics which govern open conversations (over a developing domain, where the mesh is extended) allow A to aim for and to work on and ultimately to understand nodes that have temporary names and an undetermined index.

2.7.4. It is important to separate the notion of an occasion, typifying some orientation of A to B or vice versa and an instant of time (which is indexed t_n during occasion n). The index mechanism is a clock, starting from $t_n = 0$ when occasion n begins.

2.7.5. Both instants, t, and occasions, n, are serially ordered. But, during an occasion, several processes may (and generally do) go on concurrently. A's goal is usually to learn about several topic relations which he tackles at once, whilst others are explored. These working points or "loci of control" are properly indexed by separate and potentially independent clocks (which may be locked into partial synchrony by interaction between the processes in each "locus"). In fact, occasions themselves can only be serially

ordered because of the arbitrary restriction of Section 2.7.2. (One and only one aim-for node).

2.7.6. In contrast, the activity involved in giving an L^0 explanation (a mandatory action for any node that becomes marked as understood) is serially ordered. This is a constraint imposed for mechanical convenience and it has no positive significance whatever. As a result of experimental work it looks as though many students do not furnish serial or string-like explanations, though (viewed through the screen of CASTE) they seem to do so. Moreover, the seriality restriction impairs the system's operation and the machinery is being modified to eliminate this constraint or, at least, to minimise its effects.

2.8 *Outline of the construction of a pruned cyclic entailment mesh and its L descriptors.* The pruned cyclic entailment mesh is derived in stages from a metalinguistic representation of the subject matter which is a relational network in which nodes stand for topic relations R_i (they do so throughout) and arcs stand for the relations between topic relations. It is a basic tenet of the theory that all relations between topic relations can be expressed, up to isomorphism, by some combination of relational operators drawn from a set of relational operators under which the field of relations is closed. The topic relations themselves reflect, again up to isomorphism, the topics mooted by a source, generally a subject matter expert, in a freely interpreted and natural language in which ungrammatical expressions may be acceptable (the metalanguage used to speak of operators etc. is usually a natural language also, but its grammar is precise and its interpretation is always open to justification). At this stage, only two conditions are essential; namely, that each topic cited (apart from the first) is somehow connected to some other topic and that any topic has one and only one name (a consistency requirement).

2.8.1. In the interests of uniform representation (relational operators are, themselves, relations) the relational network is transformed to yield an entailment network in which relational operators are also depicted as nodes but distinct derivation paths (for obtaining R_i from R_j for example) are preserved and in which the directed arcs represent one kind of dependency called entailment.

2.8.2. Next, the entailment network is denuded of relational operators because it is assumed, as later, that they are isomorphic images of L^1 procedures, $\underline{Proc}^1_A$, in the repertoire, π^1_A of any student. When the node representing a relational operator is deleted, the potential information lost in the process is encoded in an operator data-base together with indices of the position at which the operator was applied. If a node is marked aim-for the student can gain access to the operator data-base.

2.8.3. The resulting entailment mesh for a conversational domain in R must be *cyclic*. This means that any R_i may be reconstructed from other R_j in R using the specified relational operators. In particular this is true of R (Head). Cyclicity has two concomitants; one bearing on the subject matter and the other bearing on our theory of learning.

2.8.4. Regarding the subject matter the existence of cyclicity implies that the relations in R constitute a complete theory; an explicable point of view. The explanation cycle or the theory is a Gesalt. If a mesh, R, is cyclic then a certain relation R_i (called the head (in Section 1.8.5)) is L explicable without loss of specificity and without invoking relations other than those named in this mesh.

2.8.5. So far as the learning theory is concerned, cyclicity implies that if there are $\underline{Proc}^1_A$ in π^1_A isomorphic to the nodes representing relational operators (as assumed in Section 1.8.2, to justify deleting these nodes earlier in the construction process), then A is able to reproduce any $\underline{Proc}^0_A$ i that exists as a concept (in π^0_A) which brings about a topic relation R_i. These concepts are learnable and may be reproduced as memories in a domain R. Further, if the mesh is consistent (as required) in addition to being cyclic, and if the conversation remains in the conversational domain of R, then memories/concepts will not decay or disappear due to any kind of abrasion or interference.

2.8.6. A test for cyclicity is applied at any moment the subject matter expert feels he has stated "his subject", and points to some topic relation he deems to "head" the subject matter. It should be emphasised that a choice of head does not bear directly upon the relations which make up the conversational domain R. Rather, this choice is the first and most important step in giving an L^1 description of R. If the cyclicity test is successful, or if the mesh can be rendered cyclic by adjoining other explanations or

derivations, then it is pruned. In the pruned entailment mesh, the arcs corresponding to reproductive paths have been deleted and the mesh has been disconnected from peripheral relations surrounding it.

2.8.7. During the pruning operation it also becomes evident that certain topic relations have been regarded by the subject matter expert as *primitive*; i.e. as topic relations any student will know and look upon as properties. At this point in the analysis, an expert's usually tacit standards of the subject are made explicit. There are thus two assumptions about the student (A), underlying the postulate that A can learn the R_i in R. One (noted already) is that π^1_A contains $\underline{Proc}^1_A$ isomorphic to each relational operator represented by a node distinguished and deleted in forming the entailment mesh. The other is that π^0_A contains $\underline{Proc}^0_A$ for bringing about all the primitive topic relations.

2.8.8. In citing the head topic relation, the subject matter expert induced one descriptor (actually, in this case, "2nd"). He is required to produce others (in this case "1st" and "3rd") to satisfy the requirement that L^1 expressions in the descriptors furnish a complete indexing and pattern description system. Optionally other descriptors may be added.

2.8.9. $D^1(R)$ is a collection of L^1 statements that consist in the pruned entailment mesh, the descriptor values and an L^1 description of the operator data base. All of them are physically coded in ES(R) the entailment structure, under the correspondences already described. But ES(R) also provides a physical vehicle for expressing L^1 statements (uttered either by A or B); for example, the marker predicate statements (asserted of each R_i in R) and ostending statements that indicate place holders for relations, classes or hypotheses.

2.9. *Task description.* Data links are established between each node in ES(R) (hence, each topic relation described in $D^1(R)$) and a physical embodiment of the L^0 description, $D^0(R_i)$ of how this relation may be brought about. $D^0(R_i)$ is the L operational meaning of R_i and there is a link (bidirectional pointer) connecting the node i in ES(R), which is a place holder for topic relation R_i, with an embodiment $TS(R_i)$ of $D^0(R_i)$ as in Fig. 5. In aggregate, all of the $TS(R_i)$, related, so far, only via ES(R), constitute the task structure TS(R) which embodies $D^0(R)$. Fig. 5

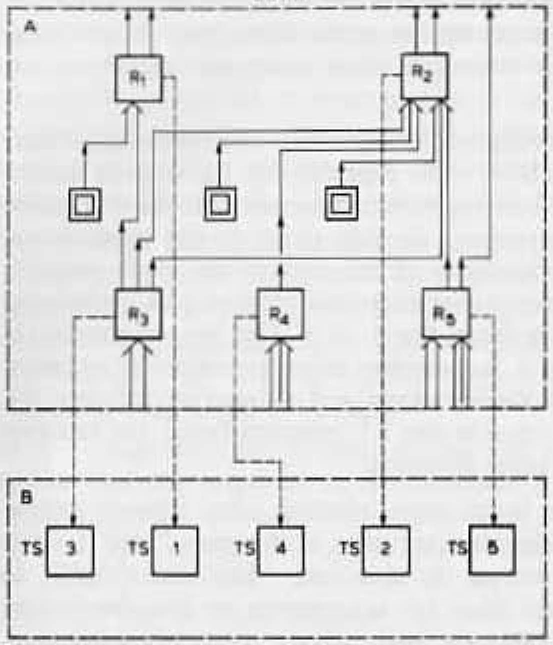


Fig. 4.5. Association between $D^1(R)$ and $D^0(R)$ or, equivalently, ES(R) and TS(R). A = a fragment of $D^1(R)$ (its concrete form is ES(R)); B = a fragment of $D^0(R)$ (its concrete form is TS(R) and is made up from $TS(R_1)$, $TS(R_2)$, $TS(R_3)$, $TS(R_4)$, $TS(R_5)$).

thus represents a small fragment of <ES(R), TS(R)>, which contains the L description $\langle D^1(R), D^0(R)\rangle$ of the conversational domain.

2.10. *Basic structures.* The pruned cyclic entailment mesh is usefully conceived as an "and/or tree" (as the phrase is used in computation science, often with reference to Boolean implicants) which has two curiosities; namely (a) Its terminal branches are "floating" to be tied down, at a particular occasion n, by the nodes of those relations R_i which, on that occasion, A is able to regard as unanalysed properties. (b) The framework is augmented by internal cycles, preserved under pruning, that represent any analogy relations holding between parts of the subject matter. (Each node in the entailment network is in a cycle, of course, and

this cycle represents an analogy relation. But the majority of the cyclic connections are deleted from the display by the pruning operation.)

Viewed thus, the pruned mesh is made up from the conjunctive, disjunctive, and analogical substructures shown in Fig. 6. Of these, the conjunctive substructure reads "node α is entailed by node β

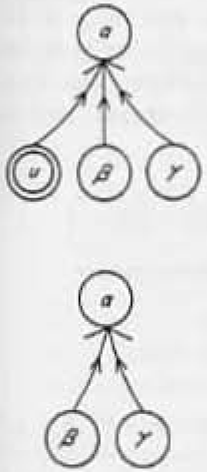


Fig. 4.6 (a). Above: Conjunctive substructure in entailment mesh; relational operator at distinguished node, U. Below: Denuded mesh meaning, "α entails (is organization of) β and γ".

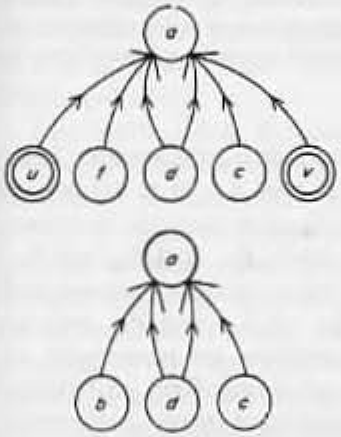


Fig. 4.6 (b). Above: Disjunctive substructure in entailment mesh; relational operators at distinguished nodes U, V. Below: Denuded mesh meaning "a entails (is organization of) b and d or of d and c or both of these".

and node γ" with the operational meaning that A is in a position to learn $\underline{Proc}^0_A \alpha$ (supposing it is not already in his repertoire) if and only if he has $\underline{Proc}^0_A \beta$ and $\underline{Proc}^0_A \gamma$ as prerequisites. The disjunctive substructure reads "node i is entailed by node a and node b or by node a and node c (equally, node d, node e or all of them, in place of the common node a) or by both of them". The operational meaning is that A is in a position to build $\underline{Proc}^0_A$ i (supposing it is not already in his repertoire) if and only if he either has $\underline{Proc}^0_A$ a and $\underline{Proc}^0_A$ b or $\underline{Proc}^0_A$ a and $\underline{Proc}^0_A$ c or both of these in his repertoire. The analogy relation substructure (Fig. 7) is a complex made up from cyclic and disjunctive forms and is

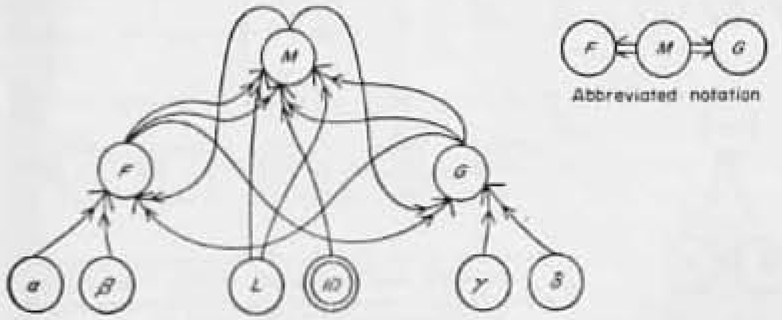


Fig. 4.7. Analogy relations of this type remain in the completed entailment structure. M = Analogy in universe, L, between functions or relations F, G, with domain, range, α, β and γ, δ. Given "objects" $\alpha_0, \beta_0, \gamma_0, \delta_0$ in domain, range, of F, G, the structure leads to simple analogical expressions, like "α_0 is to β_0 as γ_0 is to δ_0". Distinguished node stands for isomorphism (here, in universe L). Plate 3 notation on right.

usually written in an abbreviated notation. It means that R_E is an analogy between the topic relations R_F and R_G that holds in a universe Z. Given knowledge of R_E and R_F it is possible to obtain R_G. Given R_E and R_G it is possible to obtain R_F. But R_F and R_G may be known in other ways than by this analogical process and R_E may be known through knowledge of R_F and R_G and an analogical universe R_Z in which these relations are isomorphic or independent. For example, a student of elementary probability theory could learn about abstract topics and about real world topics and grasp their analogical relation. Alternatively, he could learn about abstract topics and derive analogical information about the real world, or he might learn in all these ways.

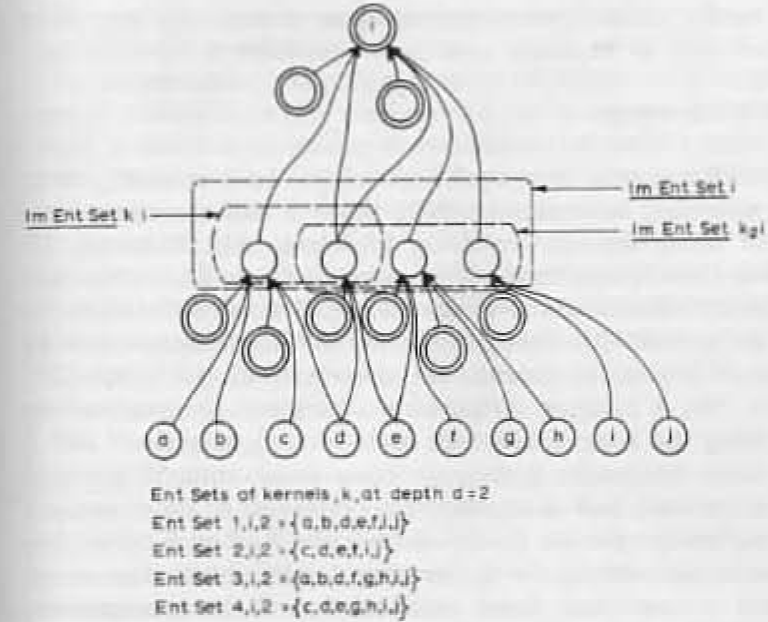


Fig. 4.8. The immediate entailment sets of a node i and a complete listing of the entailments sets at depth of 2. Ent Sets of kernels, k, at depth d = 2; Ent Set 1, i, 2 = { a, b, d, e, f, i, j }; Ent Set 2, i, 2 = { c, d, e, f, i, j }; Ent Set 3, i, 2 = { a, b, d, f, g, h, i, j }; Ent Set 4, i, 2 = { c, d, e, g, h, i, j }.

2.11. *Classes of nodes.* The following comments are repeated and detailed in later chapters. At this stage, it is convenient to have a glimpse of the classes or configurations of nodes that are illustrated in Fig. 8.

Any node i has at least one *kernel* obtained by tracing out conjunctive entailment chains and noting the nodes that are their members. A disjunctive node has several kernels; the number generally depends upon the depth, d, to which the enumeration process descends. The kernels of node i are indexed by k_i, at a specific depth, d, the maximum value of k_i is determined and the nodes in these kernels are called the entailment sets of node i denoted by Ent Set k, i. The union of the Ent Set k, i over all the kernels is called the entailment set of node i and is denoted Ent Set i. An immediate entailment set Im Ent Set k, i of node i is Ent Set k, i for the special (one removed) case of d = 1, and the union of all the Im Ent Set k, i is the immediate entailment set, denoted Im Ent Set i.

The entire pruned cyclic entailment mesh is the Ent Set of its own head node or its several analogous head nodes.

3. The CET heuristic

The CET heuristic is designed to achieve understanding of all that is necessary to comprehend R(Head). It can be expressed in terms of an L dialogue involving L^1 commands (to learn), L^1 questions (to explain how), L^0 commands (to build models that bring about relations and correspond to practical explanations) in reply to a tacit (or overtly paired) L^0 explanation-eliciting question. A's direct commands are severely restricted by the CET heuristic. But A is in a position to give indirect commands of the kind "bring the topic relation of node i to my attention" and B obeys these commands if they are (in a sense outlined below) L legal. In contrast, B is in a position to command A; for example if A has indirectly opted to learn topic relation R_i then B insists that he does so successfully. In the process A may obtain demonstrations and tutorial data about topic relation R_i by appropriately commanding and questioning B.

In particular, the experiment starts with B issuing A the L^0 command "Model R(Head)" and the corresponding L^0 question "Explain R(Head)". Unless the student knows how to explain R(Head) (if so, he is rejected as too knowledgeable), he is unable to obey the L^0 command as it is given, and interprets it, instead, as an L^1 command to learn to obey R(Head). This transition (an L^0 command that cannot be obeyed is converted into an L^1 command to learn) applies not only to R(Head) but also to other topic relations.

For the study under discussion, it was possible to ensure, as an initial condition at occasion $n = 0$, that all of the lowermost nodes are marked as being understood (and, in general, no others). It is worth emphasising, once again, that this is a very special though experimentally desirable state of affairs. The general requirement is complex but is sufficiently stated as "the CET heuristic (Appendix D and E) can be executed".

3.1. It is expeditious to describe the CET heuristic, initially, as it is seen from the student's point of view, in terms of the transaction types of Table 1. To visualise the process as the student sees it, recall that he receives a continuous display of the

values of the marker predicates (describing states of the nodes) over all nodes in ES(R).

(a) Exploration (yielding the values of further L^1 descriptors) is always allowed and (in common with all other ostending transactions) is based upon dialling the descriptor values of a node.

(b) The student is able to aim for any node he can describe, i.e. any topic relation he can appreciate provided it is not marked "understood". Appreciating a node does not imply that the student is in a position to model the corresponding topic relation or even to access and make sense of tutorial information which delineates this topic relation. These possibilities are not, however, excluded. An aim for node (only one at once in the present system) is marked aim by a white signal lamp.

(c) For any topic relation subordinate to aim (technically in the Ent Set of aim) it is possible to specify a goal, which is a list of one or more topic relations the student wishes to work on, by accessing tutorial information (such as demonstrations and descriptive data). The only further proviso is that the node is not marked as understood.

(d) The B heuristic checks the goal-statement for legality which obtains if for each node i in goal, for at least one k_i all nodes in Im Ent Set k, i are marked as understood. The B heuristic ensures legality before a goal node is established in workset by carrying out a series of recursive operations that are detailed in the computer programme of Appendix D and may involve the student in aiming for other nodes (the TagAim procedure) or furnishing mandatory explanations. Legal goal nodes are placed in workset and marked by a red signal lamp.

(e) Any node or cluster of nodes in workset can be accessed as subgoals either one at once or concurrently. If so, the node or nodes in question are marked (by intermittent red signal lamps). The student is committed to modelling the topic relation of each subgoal node and thus trying to explain it.

(f) Once a subgoal (to learn R_i) is selected, the student is requested by an L^0 command to model R_i or equivalently he is required, by an L^0 question, to explain R_i.

(g) A modelling operation is a constructive operation performed on STATLAB and described in Section 5. It may be preceded by the receipt of demonstrations (given on request) or the receipt and assimilation of tutorial data.

(h) A student submits his model for evaluation when he is satisfied that it is working, i.e. he believes it to be complete and correct and able to bring about the topic relation R_i of his current subgoal. Any model is accepted if it is evaluated complete and correct. If not, the student is referred back to demonstrations or tutorial data.

(i) If the explanatory model is accepted then the node is marked "understood" by a green signal lamp.

(j) For any occasion there must be an aim node; for any aim node there must be a goal specification; for each goal specification there must be at least one subgoal and for each subgoal at least one explanation must be attempted (albeit after tutorial information has been gathered). These conditions ensure that occasions are generated in the conversation since, at some stage, the original aim node becomes understood, and at this point, is no longer a legal aim.

(k) Originally, rather elaborate "time out" facilities were introduced to avoid slothfulness. As a matter of fact the "time out" limits have never been exceeded and will not be described.

(l) Ultimately the student receives an L^0 command/question to model/explain the original head and at this stage he is able to interpret the edict directly.

The CET control process is discussed more generally in later chapters and a detailed account of it is furnished by the annotated computer programme of Appendix D. For clarity, this programme is written with typed out statements which seem to be addressed to the student; in fact, all such statements (either input or output) are transduced by the local processor into CASTE transactions (for example, input dialling operations or modelling operations on STATLAB; outputs illuminating signal lamps or giving the indices of tutorial data files).

3.2. At the outset, an occasion counter is set to $n = 0$ and thereafter incremented by unity, for each understanding detected, until, at occasion $n = N$, the student shows understanding of the head topic (the value of N is not initially determined). An understanding in CASTE is a precise analogue for the verbal understanding reached in teachback.

The CET heuristic detects an understanding and marks the node of the current subgoal as understood if and only if

(1) the topic relation R_i at the subgoal node is correctly and completely explained.

(2) the series of aim, goal and subgoal statements, up to the current occasion, constitute an explanation of how the student learned to explain R_i.

Condition (2) is necessarily satisfied (we say why it is in the next chapters) provided that the student acts according to the rules of L and obeys any momentary additional restrictions imposed by the CET heuristic; in other words, the CET heuristic is designed to ensure that if a student does select aim and goal etc. (which he must, by edict) then a record of this behaviour is also an L^1 explanation. This record is, of course, displayed as a distribution of illuminated signal lamps (the L^1 marker predicates) over the entailment structure. It is visible to the student and is electrically represented to the CET heuristic. This record of the L^1 explanation is also called A's learning strategy; unqualified, if it extends up to occasion N, or "up to the nth occasion", if the sequence has n entries; $N > n$. Some examples of learning strategies, to be discussed in due course, are shown in Plate 6 and Plate 7. The reader may find it helpful to glance at these in order to obtain an overall picture of a learning strategy's appearance.

In equating understanding with teachback, as we did earlier, it is useful to look at the order in which these conditions are satisfied. For teachback, condition (1) is satisfied first; condition (2) is checked, by retrospection, after a correct and complete explanation has been elicited. For the straightforward operation of CASTE the order is reversed. Since a learning strategy is available, the CET heuristic "knows how the student learned to explain R_i" before it "knows whether or not he can actually explain R_i". Thus at any occasion, n, the CET heuristic is in a position to make the following conditional statement.

"If R_i is explained during this occasion, then condition (2) is satisfied", or "Condition (1) will be satisfied if condition (2) is satisfied".

Though the reversal seems, at first sight, counterintuitive, a moment's reflection will convince the reader that most everyday conversations which lead to understanding are subject to similar inversions. It simply is not always true (and it does not seem to be usually true) that "I can explain" before "I can explain how I explain (if I explain at all)". With the exception of some kinds of insight, people are commonly aware of how they will tackle a problem before they tackle it, and CASTE captures this often overlooked feature of reality.

4. Typical Data about Learning Strategies

The 20 odd frames of Plate 6 and Plate 7 show the learning strategies adopted by two different students.

For graphical convenience signal lamp markers are replaced by triangle (aim), circle (workset) and by rectangle (understood). Each frame describes one occasion, indexed n, and a sequence of frames constitutes a learning strategy. This may also be regarded as the student's L^1 account of how he came to know and thus explain R(Head). Plate 6 and Plate 7 are "typical" in the sense that they do not differ, significantly (Section 11), from the patterns of other students of the same type in respect of the number of nodes in workset and the location of aim.

Plate 6 depicts a typical Holist Learning Strategy. Aim is distinct from workset (until aim is marked as subgoal) and there are multiple goals in workset. Using subgoal statements the student garners and processes data relevant to aim from various, often widely spread out, topic relations. In general the act of learning about one subgoal bears upon learning about all the other goals in workset. All Holist Learners are true to this type.

Plate 7 depicts a typical Serialist Learning Strategy. Aim usually coalesces with the one goal in workset; progress is nested and step by step. All serialist learners are true to this type.

Holist learners and serialist learners are not primarily discriminated in terms of these patterns, although they could be. Before the experiment, the strategic tendencies of each respondent were independently determined (Section 11) with respect to entirely different subject matter (amongst other materials the taxonomies of Chapter 3). In one pretest situation the respondent is free to learn in any way and is found to adopt a holist or serialist approach (judged by a standard and quantified content analysis of protocols). In another pretest situation the respondent is taught a subject matter in a manner adapted to a holist learner which is deliberately contrived to present the serialist with difficulties; once again, each respondent is "typed" as having a disposition to learn in a serialist or holist manner (as it happens, respondents are "true to type" over both pretest situations). The empirical finding is that a respondent with holist disposition unequivocally adopts a holist learning strategy during a CASTE conversation over the probability theory domain whenever he is permitted (as he is by the CET heuristic), to learn as he likes.

Similarly a respondent typed serialist, adopts a serial learning strategy.

The phenomenon of cognitive fixity (Chapters 2 and 3) is partly responsible for the very definite holist/serialist dichotomy shown by the results in Plate 6 and Plate 7. Most likely, cognitive fixity also accounts for the clearcut pretest results.

The phenomenon itself is not demonstrated as a separable influence by an unbiased CET heuristic, because under this heuristic, there is nothing to prevent the student learning in any legal way he likes. A more convincing demonstration is furnished by other experiments in which students who were "fixed" in one mode (which did not in fact suit them) were provided with evidence indicating the superiority of another mode. A great deal of "persuasion" is required in order to make most students change modes voluntarily. The necessary amount of evidence also increased (rapidly at first, more slowly later) as the preferred habit becomes ingrained and it may even be impossible to induce a transition whilst the conversation remains anchored on the same domain.

5. Explanations and Demonstrations as Modelling Operations on STATLAB

One crucial process in the operation of a CET heuristic is an explanation, from A, in reply to a suitable question. In a context where condition 2 of Section 2.2 is satisfied, for example, a complete and correct explanation leads to the event that the node i of R_i is marked understood (Table 2); further, this is the only way in which that essential event can take place.

When STATLAB is used (as it always is in our main example) to elicit non-verbal explanations, each explanation-eliciting question is paired up with a command to make any correct and complete model that brings about R_i and A's response in the environment of STATLAB (a modelling operation) is interpreted as an image of A's explanation. Such an interpretation is assumed throughout the present section. The extremely general command (make any correct and complete model) is called a *Base* command. It may be specialised by qualifications.

5.1. *Obtaining explanations.* For the subject matter of probability theory the task structure TS(R) which embodies $D^0(R)$ is a physical laboratory like analogue, (STATLAB) for carrying out modelling operations usually by plugging up connections between sockets on the display. It is augmented by the computer programme in Appendix E which receives input from local processors attached directly to STATLAB. The equipment was built in response to the demands made by this subject matter area: it also seems that no comparable "probability laboratory" existed. But, because TS(R) is made in the metal (and for many subject matters this is the most convenient technique) it need not be explicitly formalised.

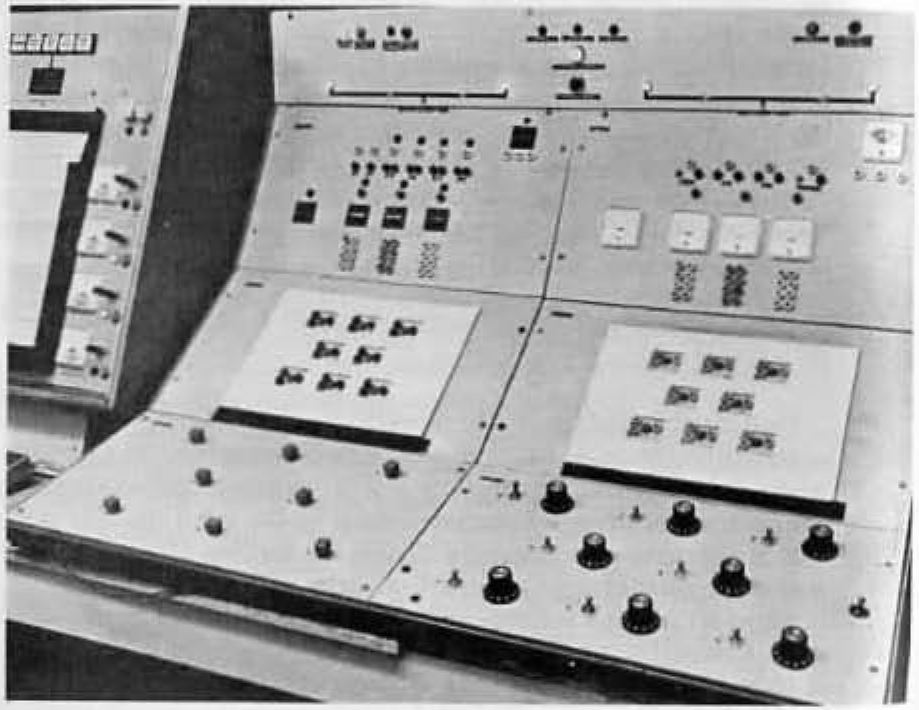


Plate 4. Close view of STATLAB. Lower left quadrant contains modelling facility for real occurrences, under student (button) control and tape control: signified by illumination of lamps and the energising of sockets. Subsets of occurrences specified by plugging into units above. Lower right quadrant contains modelling facility for an abstract universe (of "events" corresponding to "real" occurrences). Upper left quadrant: Modelling facility for frequency of logically aggregated classes of real occurrences. Upper right quadrant: Modelling facility for measures on set theoretic combinations of abstract events. Isomorphism, statistical inference and other analogy relations involve both sides of STATLAB, and qualifying statements (top row).

5.1.1. A close-up view of STATLAB is shown in Plate 4 and illustrates several features of any $D^0(R)$ and any TS(R). Since STATLAB is used to elicit explanations from A in respect of explanations demanding questions posed by B, it will, henceforward, be assumed that all plugs are removed and all sockets are free at the moment when such a question is actually posed. This (unduly restrictive) requirement ensures one necessary precondition for posing the question, namely that R_i is not already satisfied.

5.1.2. The layout of the display and response panel of STATLAB is partitioned by at least one of the L^1 descriptors, in this case by "1st". The lower left quadrant is devoted to modelling abstract logical entities; in particular, forming subsets of a set of abstract events. Each event may be assigned a measure (its "probability number") by the potentiometers on the lowermost panel and any event may be removed from the "universe of events", by a switching operation. The student can furnish explanatory answers in respect of any question posed under an R_i in the region of abstract logical models (1st = C) by plugging-up operations that form the entity required in any of many ways permitted by R_i. For example, one explanation eliciting question on R_{25} is "How do you model 2 or more inclusive events?" which is correctly answered by any one or more plugging up constructions that satisfy $R_{25} = R_i$. Correct, in this context, is used quite normally, i.e. an explanation is a correct modelling if and only if the corresponding model satisfied R_{25}. The only restriction imposed is that the domain of R_{25}, which is generally any finite set of events, is restricted to at most 8 events and the range of R_{25} is restricted to not more than 4 instantations of the set of subsets of the chosen event set. But no constraint is imposed upon the order of modelling.

Much the same is true for R_{28} where the question is "How do you model all forms of complement of composite events?"; again, the explanation is judged correct if the model submitted satisfied R_{28}. But the relation R_{28} is more complicated than R_{25}; for example, the models must demonstrate the proper (one and only one) interpretation of exclusive disjunction as well as demonstrating the concept of the complement of a composite event (subset of events) with respect to the universe of events chosen by the student.

As a final example from this quadrant the explanation eliciting question on R_{27} calls for inferences under complementation and, as a result of this, certain parts of the explanation must (due to the nature of R_{27}) be serially executed, i.e. a model cannot be correct (a member of R_{27}) unless these constraints are respected.

5.1.3. The student's plugging-up operations in reply to a question are sensed electrically and compared with a rule which is $D^0(R_i)$ restricted to domains and ranges comprehended by the instrument. The rule could be incorporated in any suitable medium (for example, in a computer programme) and it may ($D^0(R_{29})$) or may not ($D^0(R_{28})$) involve serial dependencies. As it turns out, the most convenient method of writing the rules (and mustering them for comparison with models when R_i is addressed by an explanation-demanding question) is to incorporate part of the rule in a special purpose machine and part of it in the routine listed in Appendix E.

The outcome of the comparison is either that the model does satisfy R_i or that it satisfies only part of R_i or that, in some respect or other, it fails to satisfy R_i.

The use made of this information (especially of partial mistakes) depends upon the B heuristic. But all explanation processing routines have the property that a model which partly satisfies R_i gives rise to further questions, namely "Is there any other model that does so in a different way?" until such moment as the student's constructions, in toto, instantiate all of R_i and he has thus given a correct and complete explanation. For this purpose, the L^0 explanation processor stores, temporarily, the value of previous comparisons and a list of the constraints that still have to be exhibited in order to satisfy R_i.

5.1.4. The upper right quadrant of STATLAB is devoted to measures on subsets of events given an assignment of measures (probability numbers) over certain simple events; the subject matter of 1st = H. Measures on subsets (the composite events) are displayed, by the 3 lower meters, as sums. An analogue computer is employed to realise the operators "+", "−" and "1−" that are engaged, by way of plugging up operations, to combine measures on subsets in forming a resultant measure, p, which (being a probability number) must satisfy $1 \geqslant p \geqslant 0$. The upper right hand quadrant also involves the notion of "conditional probability number" in the sense of the measure resulting if a statement is made about the measures on certain of the subsets of events; such statements being visually displayed. For evaluation the digitised values of the analogue computer variables are presented to the programme of Appendix E. Sometimes (R_{30}) the comparison involves only numerical boundary conditions (measures are non-negative fractions). Usually it also involves modelling a structure as well as satisfying numerical constraints (for example, on R_{31} the student's L^0 explanation must image the proper composition of measures over inclusive or exclusive subsets). In either case the routine in Appendix E demands structural data corresponding to the arithmetic operations that are being modelled. It may also involve a contingent statement; for example, the (conditional probability) explanation of R_{34}. But the explanation processing routine is essentially the same as it was for the lower right quadrant.

5.1.5. The lower left quadrant of STATLAB (corresponding to all of the 1st = C topic relations) is the representation of a set of real experimental results which occupy a unit interval and occur one at once (the lowermost output sockets). Results occur either at the student's discretion (the buttons associated with the result sockets) or, at the student's request, from a quasi random input tape. Subsets of results are formed by logical union and are represented by the 3 middle sockets each of which is associated with a resettable result-frequency counter. As before, the plugging up of sockets is electrically sensed and model comparison (for rule $D^0(R_i)$) is carried out by the programme in Appendix E, given the sensed values.

The upper left quadrant of STATLAB (in register with the 1st = D topic relations) provides facilities for combining subsets of results via the (plugged up) logical operators "Λ", "not", and "V" to form an overall result that has a frequency of occurrence counted on the right-most counter. $D^0(R_i)$ is still a rule, and the models submitted are compared with this rule.

5.1.6. The single topic relation R_{14} = R involves the concept of isomorphism, modelled by matching an entire model on the lower left quadrant against a model in the lower right quadrant of STATLAB. The possible universe of events (lower right) and of results (lower left) can be placed, by the student, in any one to one correspondence thereby constructing real experimental structures that are isomorphic with abstract models. For example, simple events may be permuted with respect to simple results since the sets are initially unordered. But the routine in Appendix E keeps a record, over each occasion, so that all subsequent modelling must reflect this initial assignment.

R_{14} illustrates a fairly common difficulty in its purest form. An isomorphism between the simple event sockets (set) and the simple results sockets (set) is literally any one to one correspondence. However, inspection of Plate 4 shows that these sockets are labelled by numerals so that, in fact, the student models an isomorphism by some specific correspondence. In reply, the B heuristic must continue to make enquiries of the type "Any others like it?" until all possible permutations have been exhausted; and this is true even under the convention that the sets in question are "quasi unordered" i.e. to be regarded "as though unordered" (since the numerals stand for tags, not real numbers). The finite but lengthy enquiry sequence is avoided as follows: In order to address this node, the student must already understand the notion of a permutation. Hence, only one enquiry need be made after an initial model is erected; namely, "will any permutation of the domain and the range of the isomorphism serve as well?". Yes, any; but nothing else (i.e. no many to one or many to many or one to many mappings). The enquiry is made and the reply is elicited using the signal lamps and function buttons at the top of STATLAB. In general, it is permissible to make such enquiries in respect of the field of any relational operator (permutation of domain/range coordinates is one relational operator) and to elicit the (partially quantified) replies "any" and "all" and "no other", either verbally or by way of special buttons.

5.1.7. All topic relations in 1st = F involve matching between models constructed in the upper left and the upper right quadrants of the apparatus and these matches are contingent upon models of abstract entities (built in the lower right quadrant), and models of real experiments (built in the lower left quadrant). For example the correct reply under R_{21} involves building models for experiments; making valid (probabilistic) inferences about the results and, vice versa, making valid models for quasi random experiments using a tape simulated input of experimental results. In any case, "probability", in contrast to both "frequency of results" and "probability number" is an analogy relation between frequencies and measures; i.e. the relation called valid probabilistic inference.

5.1.8. In summary, an L^0 explanation consists in a sequence of operations (plugging up, potentiometer setting and so on) as a result of which the student makes a model on STATLAB which he believes able to bring about a topic relation R_i. At any stage in the modelling process the student can pause without commitment and make certain that the model does operate in a satisfactory fashion when it is executed. Depending upon the model, the idea of "execution" may be trivial or not. At one extreme it amounts to no more than pressing result buttons and observing whether or not result lamps are illuminated in the desired subsets (if not, the model may be remade by further plugging up operations). At the other extreme execution is far from trivial; for example, running a quasi random "experimental results" tape as input to a model for an experimental design and determining the outcome frequencies produced.

The rule, in either case, is that execution does not take place whilst modelling is in progress, or vice versa, that the model is not altered if it is being executed as a dynamic entity. If these conditions are satisfied, the student can make and remake a model until he is satisfied with its performance. He is only committed to the model at the point when he is satisfied (though his hypothesis that the model is correct and complete may or may not be verified) and the model is submitted for evaluation as correct and complete or in some manner defective. Submission is signalled by pressing a special button on the control panel.

5.2. *Demonstration.* STATLAB can be operated in the converse manner in order to demonstrate a relation R_i; that is, B gives an explanation of this relation to A. When used for this purpose STATLAB is augmented by overlay cards (Plate 5, Plate 6) that are manually withdrawn from the tutorial data file and placed on the panels to label sockets, buttons, meters, counters and signal lamps. Some explanatory data is inscribed on each overlay card (for example, plugging up instructions and odd pieces of advice): other information is obtained from the tutorial data file at the same index position as the overlay card, either manually or by projection (when a slide is indexed on the random access projector). Formally, all of this is an L statement from B to A about their common environment; namely, STATLAB and its peripheral equipment. Demonstration also involves one other component: namely, one or more B operation upon the environment. For example, if the demonstration of R_i involves experi-

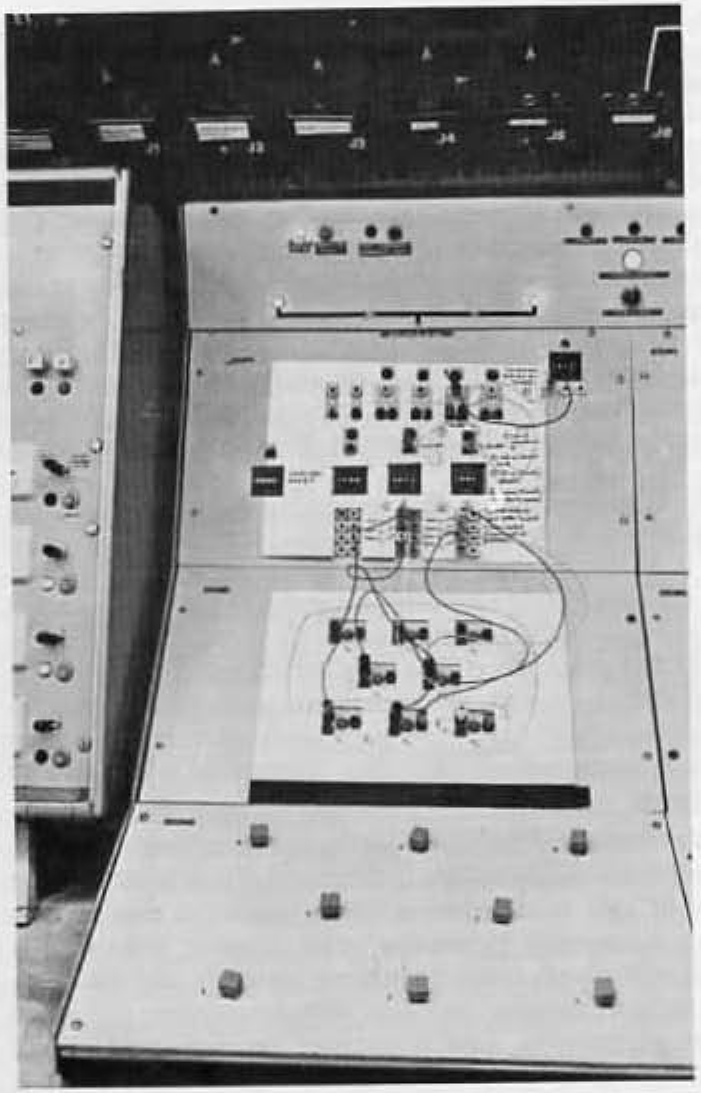


Plate 5. Layover card for demonstration (in this case using only left lower and left upper quadrant).

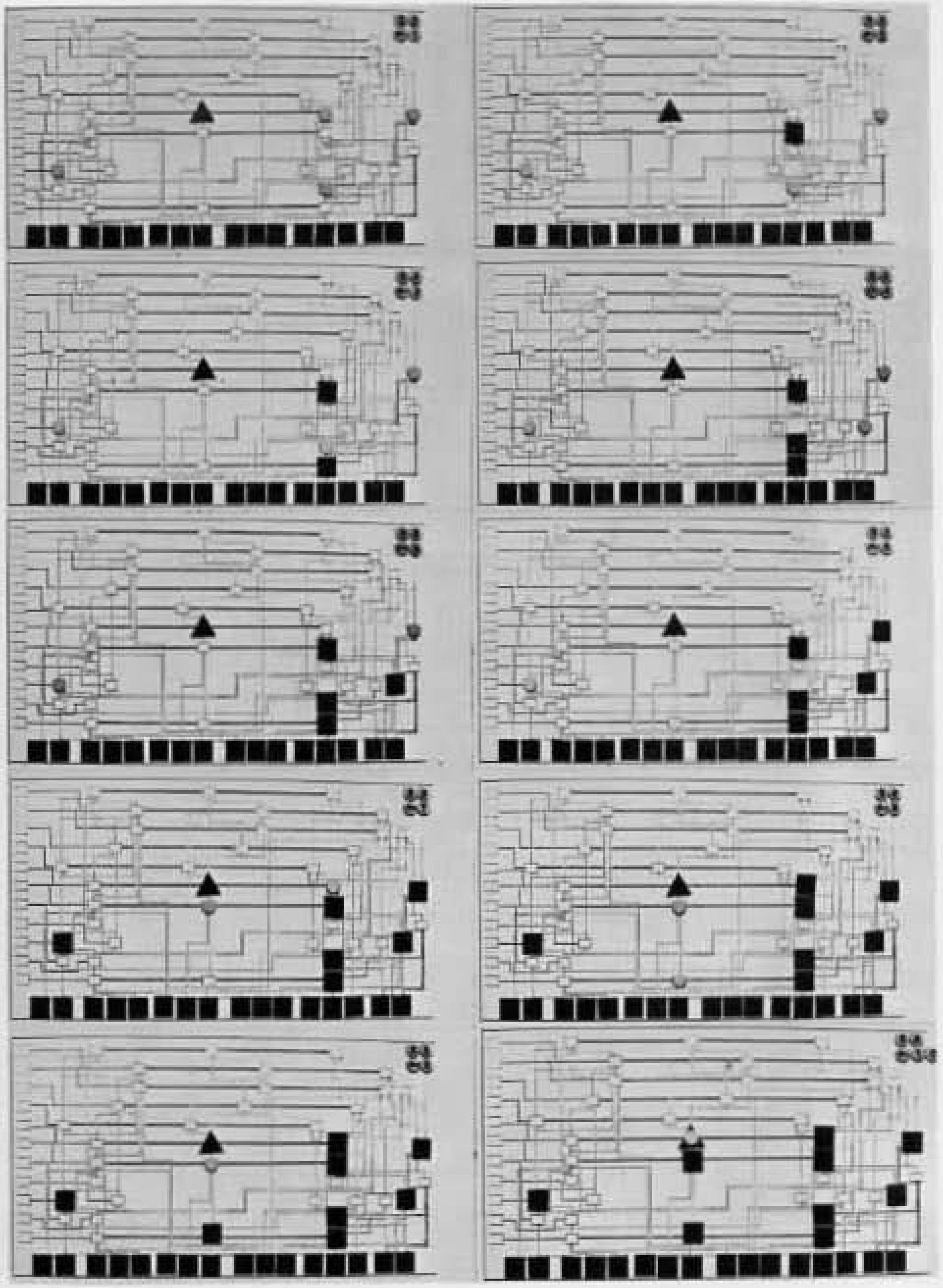
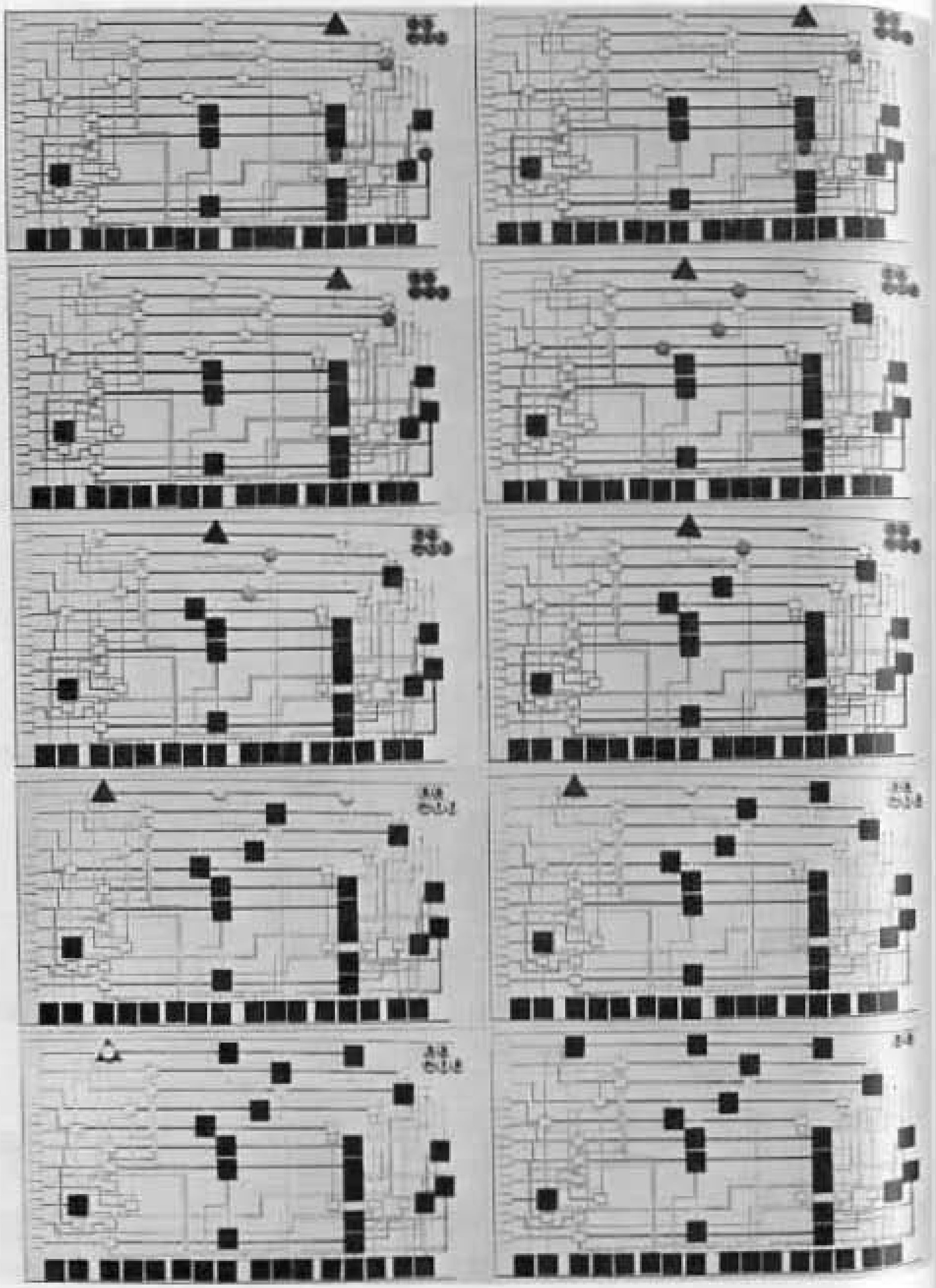


Plate 6, legend see p. 119.



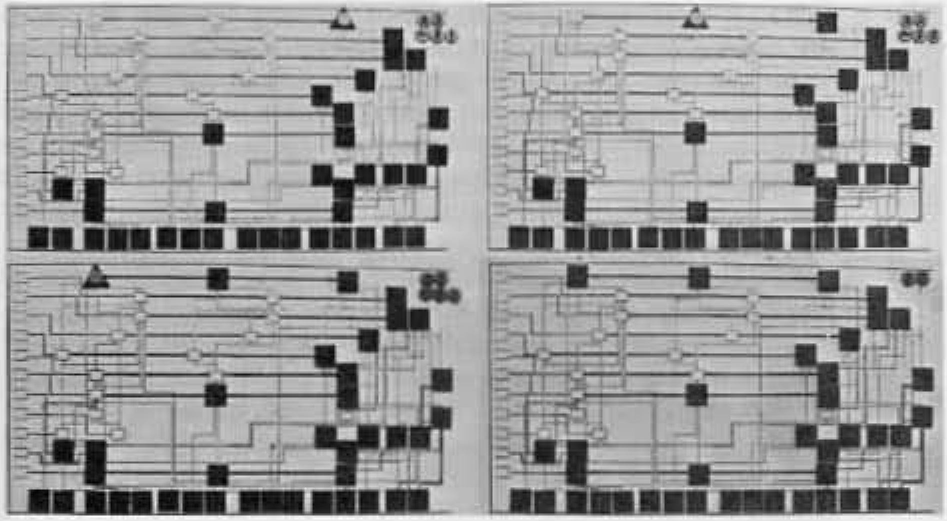


Plate 6. Learning strategy of one holist student, depicted as series of 20 frames overlying entailment structure; each frame representing an occasion. Triangular marker is aim; circular markers signify goals in workset; rectangular marker is understood.

mental results then a quasi-random tape input replaces the (student controlled) result buttons and this tape is selected by B. If the demonstration involves any real or inferred structure, then a preprogrammed model is literally selected by B (in other words, B, given R_i, selects a small machine which is fabricated by switching operations, the converse of modelling operations, which, for the rest of this particular demonstration, constrains the STATLAB environment).

Over and above its conventional and tutorial purpose, an explanation plays other parts essential to a conversation. It establishes the relations incorporated in the modelling facility (STATLAB), that are essential if it is to act as a modelling facility but which necessarily, also, restrict the class of models that can be built. In this connection, it is essential that before a student is asked to explain or model a relation he must have received a demonstration of the pertinent part of STATLAB and this condition is guaranteed by any B heuristic.

Hence, apart from its tutorial content, a demonstration is a vehicle for making some basic L statements from B to A.

5.2.1. Some L statements ostend possibly relevant properties (or relations to-be-regarded as properties) of the environment under relation R_i. These properties are identified with L^0 predicates and the states described in these terms belong to a set; later called $X = X_B$. Conversely, if the student submits an explanation, the states of his model are any sufficient (to make the explanation complete) subset of X, later called X_A. It should be emphasised that correctness and completeness depend upon agreement between A and B. Correctness is a relative, not an absolute, term and amongst other things a demonstration presents B's interpretation of the modelling facility to A, in terms of L^0 predicates. Any L^0 predicate of the environment, relevant under R_i, is a relation (to be regarded as a property) indexed by a node in the Ent Set i and either is, or is reducible to, relations indexed by nodes marked as primitive.

The data identifying these relations is inscribed on an overlay card (Plate 5); for example, the naming of certain sockets as result sockets, other sockets as event sockets, the naming of counters, buttons, meters and indicators.

5.2.2. Other L statements assert that certain noticeable features of the environment are irrelevant, in the sense that in any communication with B they must be disregarded (there is otherwise a risk of falsifying a message). Some features of the environment are so obviously irrelevant that they are not actually cited (for example, the panel's surface texture and the nebulous "mid positions" of the buttons which are actually toggle loaded microswitches chosen to give the "feel" as well as the actuality of an on/off action).

Other negated features are far from trivial. For example, the set of events and the set of results are both, in fact, ordered sets.

No attempt is made to conceal this ordering by perceptual trickery. Nor (as noted earlier) is any essay in this direction likely to succeed for, by hypothesis, all knowables (including a fortiori, all perceivables) are relations and thus induce orderings. Nevertheless it is legitimate to assert that any relation whatsoever between elements in the field of a well defined relation is to be regarded as irrelevant by A and, in this sense negated. As a result, the (actually ordered) set is regarded as unordered[3].

5.3. *More specific commands and questions.* The base command (to bring about topic relation R_i or to model it) is the most general command that can be issued in the context of the topic relation. Specific commands, or questions referenced by the base command, either call for specified types of model/explanation or (a different matter) for models/explanations that are constructed to bring about R_i only for restricted preconditions. They are of less importance than usual for a subject matter like probability theory which has a mathematical background, but they exist. For example, some of the base commands encompass other base commands that feature as specific commands under the (first mentioned) base and many of the demonstrations are addressed to specific parts of a base command rather than the command itself.

Two points are logically essential, though they will not be stressed until later.

(a) "Specific" is used in two different senses which may be at odds; a specific method of doing something does it completely; to do a specific aspect of something (for specific preconditions), whether done by a specific method or not, does only part of the original thing.

(b) It does not follow that there is one unique operation of which the others are part (which, on execution, "does everything"). Thus $D^0(R)$ is the union of several $D^0(R_i)$ and not their nesting; the several R_i in R may have quite different fields. The

[3] So, strictly, B ostends (a) the eight socket groups numbered 1 to 8 on the left lower panel of STATLAB; (b) their spatial arrangement as shown in Plate 4 and Plate 5; (c) their one to one relation to the eight buttons; (d) their one to one relation to the eight lamps. All of these (and many more besides) are relations for which, B assumes, A may have concepts. B asserts of this ostended and indefinitely large clutch of relations; that only two, namely (c) and (d) are to be regarded as relevant; all others are to be regarded as irrelevant, except (a) is transformed into a numerical labelling (the 8-adic relation "alternative set") by applying any permutation to the order of integers which these numerals commonly denote.
It does not matter if A already has the concept of an alternative set; it does matter that he has the concepts "order" and "permutation" (which, notice, are both marked as primitive in ES(R)). Hence, by an apparently devious route, B ostends and affirms the relations between the "as-though-unordered" alternative set of sockets with buttons/lamps attached and the value "real experiment" of the (primitive in ES(R)) relation between physical occurrences and their abstractions. Henceforward, the field of this relation is known by the property of "being the result set" in STATLAB.

universe of discourse or modelling is not usually homogeneous and it is not assumed to be.

5.4. *Commands, questions, tasks and subtasks.* One cannonical form of $D^0(R_i)$ is a "command graph" representing how all logically possible commands bearing upon R_i can be executed. Base commands/questions are issued, insofar as the entire graph is given an imperative interpretation (either "A bring about R_i"; a command/explanation-demanding question to elicit a model, or "B bring about R_i"; namely, a demonstration). The most general imperative/interrogative L statement with respect of R_i is a base command (and the associated explanation eliciting question). More specific commands give an imperative interpretation to sub-graphs of the command graph.

It is usually most convenient to regard $D^0(R_i)$ as depicting a task class referencing more specific subtasks (like A making a model or B giving a demonstration to student A) and to regard $D^0(R)$ as a collection of possibly unconnected tasks; they may only be related by the entailments expressed in $D^1(R)$. For example, the abstract and the real life subtasks in elementary probability theory, are related only by entailment until a student has the concept of statistical inference and is able to use it. In contrast, all abstract structural subtasks are nested together and so are all real life structural subtasks.

Hence $D^0(R_i)$ and $D^0(R)$ are called "task structures" $TS(R_i)$ and $TS(R)$. Inspection of Appendix E shows that the operating programme is written with respect of a "task structure" on the basis of which either A or B is given more or less specific subtasks to perform (i.e. to model R_i or to demonstrate R_i; to model R_i by a particular method or to give a method specific demonstration of R_i; to model R_i for some preconditions only or to demonstrate R_i for these preconditions only).

5.5. *Evaluation of the student's model.* Any model submitted by A is evaluated as correct and complete or correct and incomplete or as somehow inadequate. The evaluation is simply a comparison, under the command or question (build a model) issued to A, between A's model and one of the many putative demonstrations (or explanations) that B can produce by executing the appropriate part of $D^0(R_i)$. In this case B's execution is private, B does not show the results to A as B would do in giving a demonstration. The criterion employed for evaluation is detailed

in the computer program of Appendix E but is (in outline) as follows: If the execution of A's model matches, one to one, the execution of any of the models produced by B then it is correct and complete and R_i is marked understood; failing which, if it matches, one to one, any of the models that refer to a specific set of preconditions only, then it is correct and incomplete (and A is informed). Failing that, it is deemed mistaken and A is informed of how his model is at fault (and, for some B heuristics, of how the fault could be remedied).

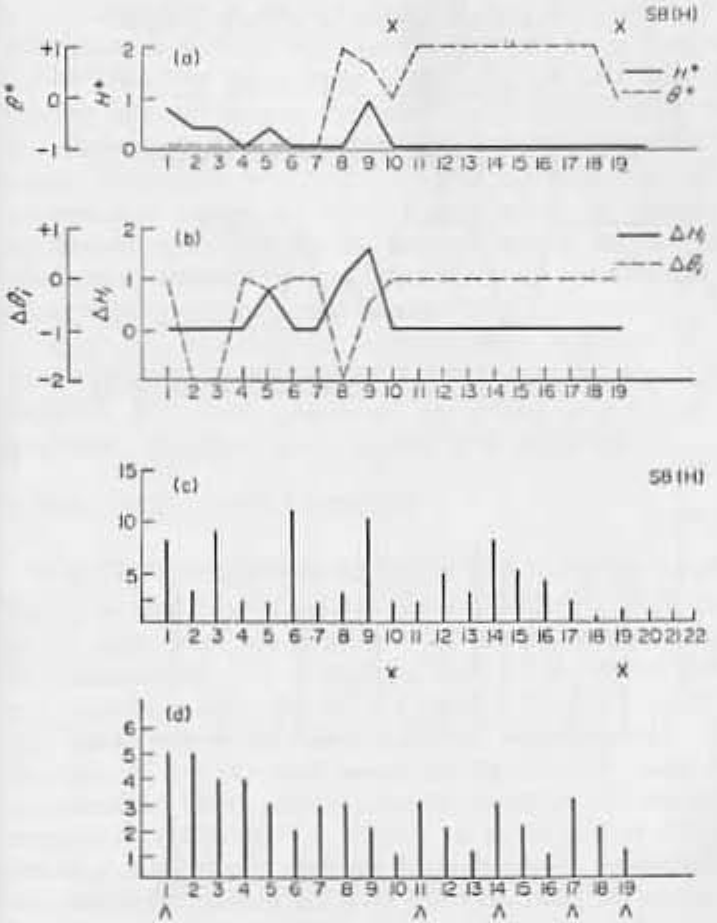


Fig. 4.9. Gross data (Plate 6). Horizontal axes show occasion No. Vertical axes as follows: (a) θ^* and H^*; (b) $\Delta\theta_i$ and ΔH_i; (c) values of J; (d) size of working on set.

6. Uncertainly Measures and Strategy Types

Although the requisite procedures have not yet been described (they are discussed in Section 7) it is usual to measure indices of problematic uncertainty (Chapter 2, Section 4) during the CET conversation even though these indices are not used directly as control variables ("problematic uncertainty" is a general name for A's uncertainty, given a problem, regarding its solution; or A's

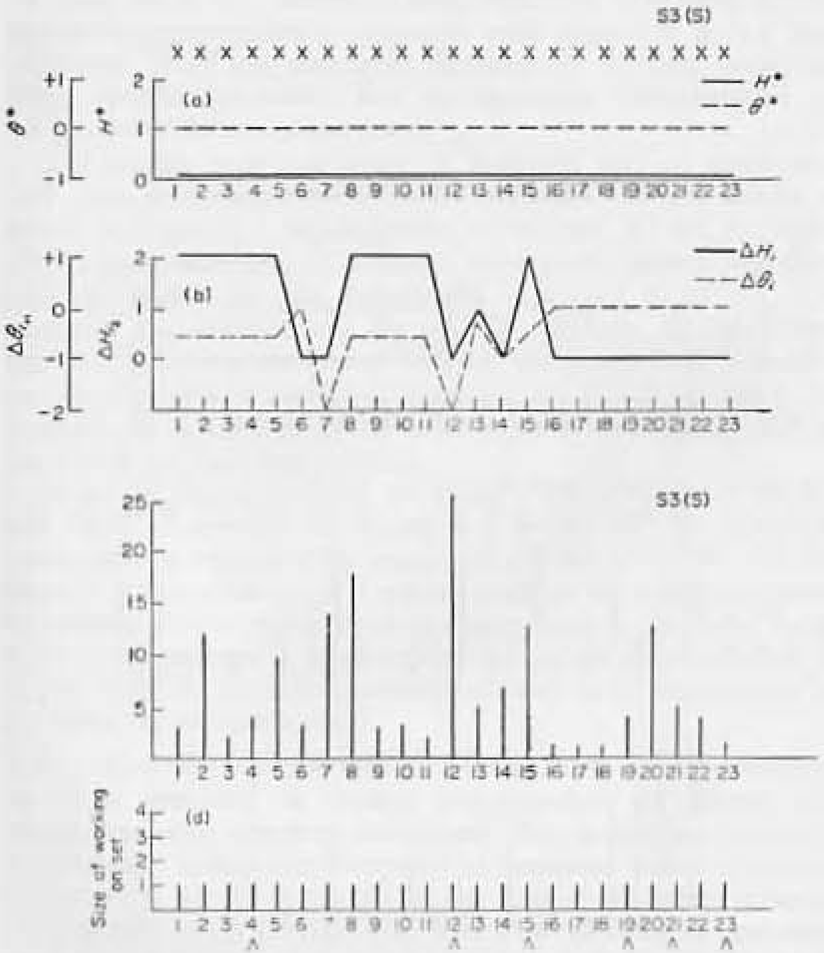


Fig. 4.10. Gross data (Plate 7). Horizontal axes show occasion No. Vertical axes as follows: (a) θ^* and H^*; (b) $\Delta\theta_i$ and $\pm\Delta H_i$; (c) values of J; (d) size of working on set.

uncertainty, given a command/question, about how to obey/answer it). An uncertainty (H) and a correct belief score (θ) are determined for each topic relation worked on as a subgoal and a distinguished uncertainty (H^*) and correct belief (θ^*) (called a "look ahead" uncertainty and correct belief) are determined for each aim node.

These values are tabulated, along with the effort (τ) and the number of nodes in the working set for the typical holist (Fig. 9 in register with Plate 6) and the typical serialist (Fig. 10 in register with Plate 7). Notably, the serialist has no uncertainty H^* or correct belief θ^* in respect of an aim node in advance of the topic relation he is working on. Serialist students do, in fact, appreciate topics ahead of those they understand (as we have found by probing in a pre-programmed fashion for correct belief measures). But the region in advance of the current topic relation where correct beliefs are entertained is canalized and often rather short; beyond that region, no belief is entertained. In contrast, a holist student is quite willing to express beliefs about many topic relations in advance of those that currently concern him and, very often, these are also correct beliefs.

Succinctly, the serialist proceeds from certainty to certainty: the holist also achieves certainty at the points where explanation is required. But these points are embedded in a nexus of dimly perceived, but often correctly perceived, relations.

7. Macro Crain Control Variables

The CET heuristic is based almost (but not quite) exclusively on the molecular or micro-grain model (Chapter 2, Sections 3 and 5) but it does, even so, guarantee the macro or molar condition of self organisation. It is possible to add a fine control over learning and cognition under the CET heuristic, by locally optimising the self organisation (of the molar or macro-model, Chapter 2, Sections 3 and 4) and satisfying various boundary conditions expressed in terms of macro grain variables. These variables are subjective probabilities or degrees of belief and are obtained from confidence estimates over sets of alternatives presented at various test intervals interleaved with the execution of a basically CET heuristic. Degrees of belief are usually aggregated by calculating information measures (H) to estimate the student's uncertainty

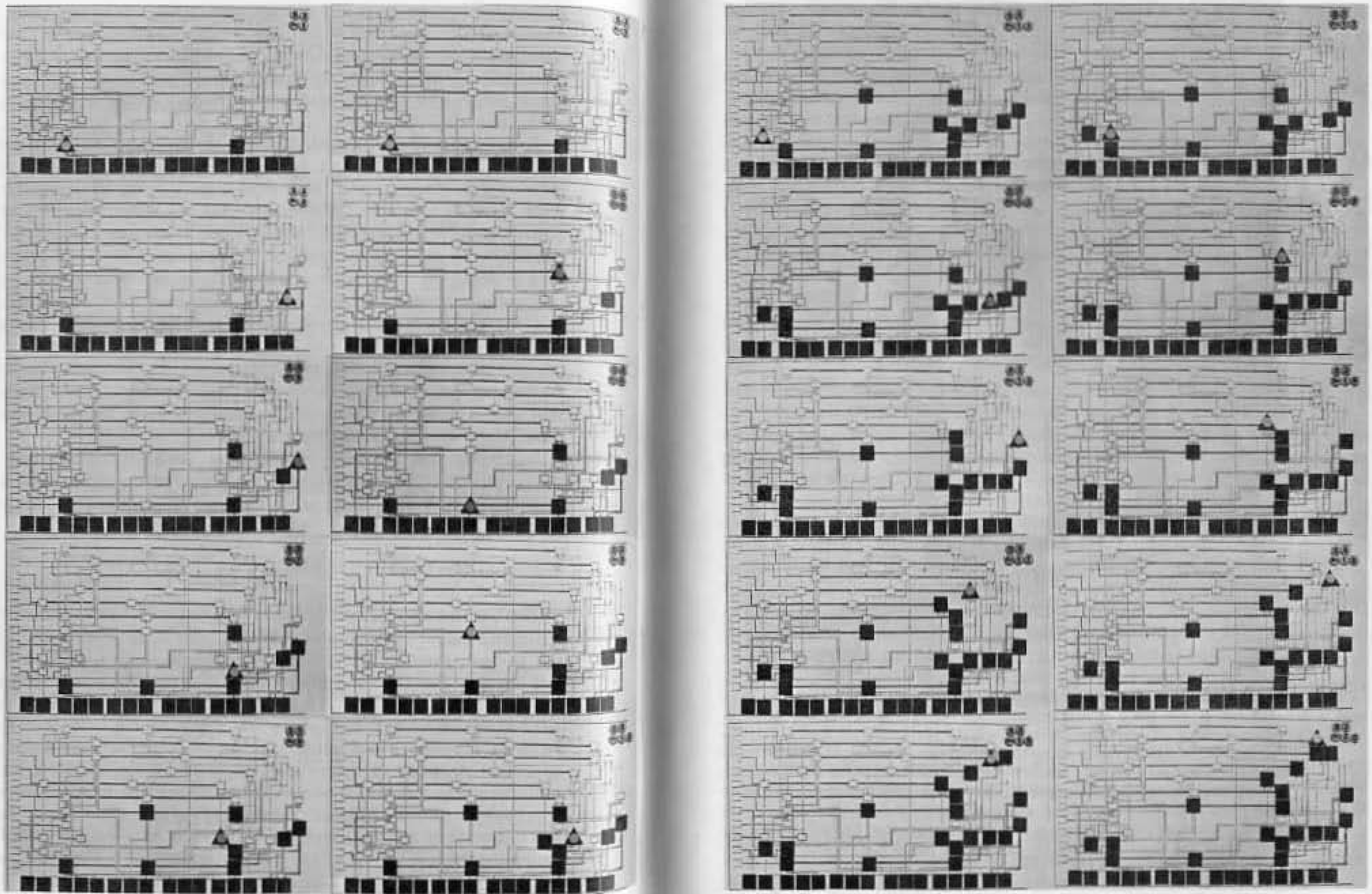


Plate 7. Learning strategy of one serialist student, depicted as series of 24 frames, overlying the entailment structure, each frame representing an occasion. As before, triangular marker is aim: the circular marker signifies a goal in workset; rectangular markers signify topic relations are understood.

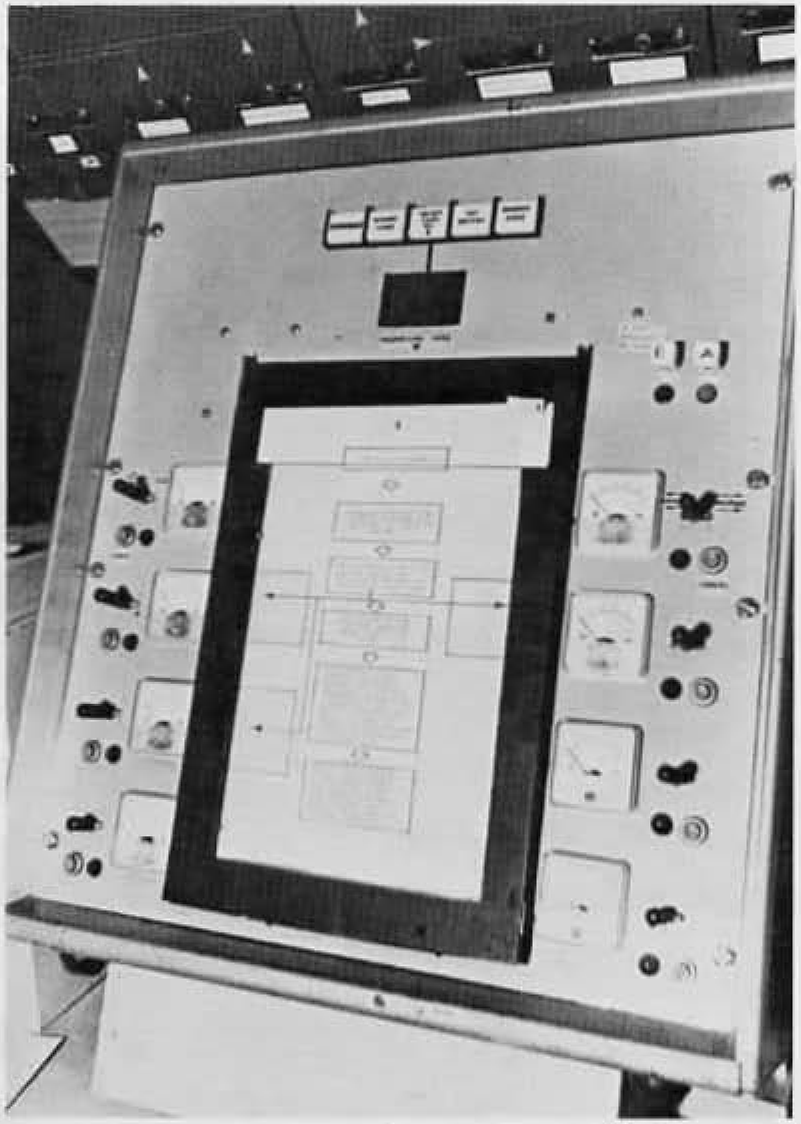


Plate 8. The Belief and Opinion Sampling System: BOSS (used to elicit confidence estimates). Digital indicators above card reader show card to be inserted. Card of test alternatives is shown positioned in the card reader (photocells in configuration along the lower edge of reader and at the edges read punched-hole codes). "Active" lamps illuminate for (up to 8) relevant alternatives on the card. Response is produced by touching centre zero levers and displayed as a normalised distribution on meters attached to each relevant alternative. If desired, respondent may cancel alternatives as "not considered" by pressing "cancel buttons". The operating sequence lamps are shown along the upper edge of the unit and the "submit response signal" is given (when the respondent is satisfied with confidence estimate) by pressing button at side.

and Shuford Scores (θ) to estimate his correct belief (Shuford et al., 1966; Baker, 1969), as noted in Section 6.

The equipment used for this purpose is the belief and opinion sampling system (BOSS); a close-up view of it is shown in Plate 8.

Questions are of the multiple-choice type and are inscribed on question cards (for any number up to eight alternatives) that are extracted from the tutorial data file (as demanded by the numerical indicators on BOSS) and inserted into the card-reader from which the codes, attached to each card, are read into CASTE. The student's response is a pattern of readings on the meters associated with each alternative that is relevant (for which a signal lamp is illuminated) and the readings are automatically normalised to yield probabilities. The student's response is made either by pressing buttons to exclude certain alternatives because they are considered irrelevant or by adjusting the levels (right for an increase in reading, left for a decrease). Initially (for m alternatives), all meters read 1/m and after an arbitrary series of adjustments a pattern is built up, which the student submits to CASTE, when he deems it satisfactory, by pressing the submit button on BOSS. The equipment also contains a rather complicated balancing circuit which simulates, as a delay, the changes in belief needed to influence the response pattern.

It is easy to construct the multiple choice questions in a standard form, since relations are subsets of a product set of their co-ordinates. The correct alternative is positioned in the subset which determines the relation (i.e. it satisfies the topic relation); the mistaken alternatives are located at a fixed signal distance from it. H (uncertainty) and θ (correct belief) are computed automatically. A BOSS estimate is obtainable, in principle, for any node in the entailment structure because each node corresponds to a topic relation under which problems may be posed by standard multiple-choice questions. The derived macro variables, the uncertainties and correct beliefs, are thus called problematic uncertainties.

If computed in respect of nodes in the workset that are selected as subgoals, the problematic macro variables of Section 6 are H (an uncertainty) and θ (a correct belief). After a subgoal topic relation has been explained, $\theta = 1$ (its maximum) and $H = 0$ (of necessity). Further, in a tutorial conversation heuristic, this condition is checked, by posing appropriate multiple-choice questions, before an explanation is requested. Hence, the initial values of H and θ on subgoal topic relations are recorded as differences; $\Delta H = H$ (Initial) $- 0$ and $\Delta\theta = 1 - \theta$ (Initial).

Problematic macro variables computed in respect of the other chosen aim node are called look-ahead uncertainties or correct beliefs (Section 6) and are symbolized as H^* and θ^*. Similar values can be sampled, in principle, for any topic relation and are denoted H_0^* and θ_0^*. But, apart from the aim node the values may be imaginary (becoming real only later in the learning process) in the sense that a student fails to appreciate the domain of the topic relation, under which a problem is posed. The act of pressing all the cancel buttons on BOSS is interpreted as a statement that the problem is incomprehensible, in which case H^* is assigned its (maximum) value, H^* (head node). If the aim node is also, uniquely, the workset node, then H^*, θ^* are numerically equal to H (initial) and θ (initial) but are improperly defined as look ahead uncertainties. In fact (as in Fig. 9 and Fig. 10) there is no look ahead uncertainty; H^*, θ^* (but not H_0^*, θ_0^*) are set to zero.

Apart from problematic uncertainty a student also experiences strategic uncertainty regarding which plan for learning he should adopt. By mandate, strategic uncertainty is resolved when one aim node is chosen, but the resolution may either be rational (the student is convinced about the rectitude of his selection) or spurious (the selection is a guess). If and only if strategic uncertainty is rationally resolved it is possible to regard H_0^* as an index of the maximum uncertainty in respect of the student's current field of attention, on a par with I^* (Chapter 2).

8. The Uncertainty Regulation (Tutorial Conversation) Heuristic

If strategic uncertainty is eliminated, it is possible to consider local self-organisation as suggested in Chapter 2. First assume that H^* (head) $\equiv I^*$ max, $H_0^* \equiv I^*$, $H \equiv I$. If so, (1) A tutorial system is self organising if either (for a given value of H) $\Delta H^* > 0$ or (for a given value of H_0^*), $0 > \Delta H$; (2) The self-organisation represents relevant learning if for any negative ΔH; $\Delta\theta > 0$. These conditions are approximated by the CET heuristic, but there is no guarantee that the rate of organisation is even locally maximised.

The question is, how should strategic uncertainty be eliminated from the picture (given which it is possible to construct heuristics for maximising the rate of decrease in relevant uncertainty with respect of the head node). Insofar as our strategy type hypothesis is correct (it is not claimed to be comprehensive since strategic types other than holist and serialist are known to exist) the crux of the matter depends upon identifying the student's strategic type and his competence to execute his chosen strategy. The conversational heuristic, on its own, ensures that a strategy is chosen by any student. As scrutiny of Plate 6, Plate 7, Fig. 8 and Fig. 10 will indicate, the chosen strategy (though not the student's competence) can be estimated by observing a short stretch of learning, no more than half an hour's worth. Alternatively, students may be pretested for strategy and competence. In the sequel, it is assumed that the chosen strategy is determined, though not the competence, at this stage.

Let us also make the following realistic assumptions in order to outline the uncertainty regulation heuristic.

(a) For $n > n_0$ there exists a set U(n) of nodes marked understood such that on tracing out the tree-like kernels of the head node there is at least one for which the branch nodes are members of U(n); the following argument applies only for $n \geqslant n_0$.

(b) That there is mean value data (examples of it are shown in Chapter 10, Fig. 1, 2, and 3) delineating the expected values of ΔH, $\Delta\theta$, τ, (effort) for each node in the entailment structure; these values are designated $\Delta\bar{H}$, $\Delta\bar{\theta}$ and $\bar{\tau}$.

Suppose the student is currently believed to be a serialist. If so, his chosen aim will nearly always also be his chosen subgoal. Under these circumstances the following heuristic is competent and is executed for each change of occasion n.

(1) Find all kernels of head that are trees with all branch nodes at some depth d from head belonging to U(n).

(2) Suppose there are P(n) of them; index them by $p = 1, \ldots . P(n)$ and call them C_p.

(3) Starting with the head node as origin, compute the value of an estimated correct organisation rate for each tree.

Thus, we wish (since aim is coalescent with subgoal) to estimate

$$\sum_{\text{nodes in } C_p} \frac{G(\Delta H, \Delta\theta)}{\tau}$$

Where G increases with decreasing uncertainty and increasing correct belief. To estimate this quantity the heuristic uses the function

$$F_p(n) = \sum_{\text{nodes in } C_p} \frac{G(\Delta\bar{H}, \Delta\bar{\theta})}{\bar{\tau}}$$

(4) Find the tree (or subset of trees) that have the maximum (or the maximal) values of $F_p(n)$; call this tree (or these trees) $C_0(n)$.

(5) If aim is in $C_0(n)$ make no comment; otherwise indicate the fact and display $C_0(n)$ to the student.

A choice on the student's part is still permitted at this step, since the CET heuristic guarantees that if the chosen aim becomes subgoal it is legal (i.e. in workset). The choice could be resolved by selecting any node g in $C_0(n)$, such that for at least one value of k all nodes in Im Ent Set, k, g, belong to U(n).

(6) If aim does not become subgoal reallocate the student's strategic type for occasion n + 1.

(7) If aim does become subgoal, the student will eventually (under the CET heuristic) explain the pertinent topic relation; at which point, node g is marked as understood, $n \to n+1$, actual values of ΔH, $\Delta\theta$, and τ are determined for node g.

(8) If $\dfrac{G(\Delta\bar{H}_g, \Delta\bar{\theta}_g)}{\bar{\tau}_g}$ exceeds $\dfrac{G(\Delta H_g, \Delta\theta_g)}{\tau_g}$ by more than an arbitrary amount (say one standard deviation unit) then recommend that the student changes strategy at occasion n + 1; otherwise, regard the information as evidence that he is a competent serialist and do nothing.

Suppose the student is classified as a holist. In general, the aim node is not in workset, though there are occasions upon which aim itself is understood before it is changed by the student (at this point, of course, the CET heuristic, on its own, renders change of aim mandatory). For brevity, call these transient occasions, and the node in question a transient node.

(1*) Suppose the chosen aim node on occasion n is node g.

(2*) If g is not transient and if g becomes subgoal on occasion n then reallocate the student's strategy type on occasion n + 1; otherwise proceed.

(3*) Find all nodes in Ent Set g that are not marked as understood by the CET heuristic; find all P(n) of the C_p (as in stage (1) above) and form the union of all nodes in any C_p that are also nodes belonging to Ent Set g; index these nodes q.

(4*) Whilst aim = g record, for each occasion, the values of ΔH, $\Delta\theta$ and τ (when the nodes in question become subgoals, and are later marked understood).

(5*) On the occasion n + r; r = 1,2,... when the aim is changed, compute the functions

$$E_1^g = \sum_q \frac{G(\Delta H_q, \Delta\theta_q)}{\tau_q}$$

and

$$E_2^g = \sum_q \frac{G(\Delta\bar{H}_q, \Delta\bar{\theta}_q)}{\bar{\tau}_q}$$

If E_2^g exceeds E_1^g by more than an arbitrary amount, recommend that the student changes his strategy from occasion n + r; otherwise proceed on the assumption that the student is a competent holist. (Recall that the aim can be changed, on any occasion, by the student. A transient occasion is not necessarily involved.)

(6*) Ent Set g has a maximal arc distance chain linking aim (here node g) to a node marked understood. Let this maximal distance be Q(n). Thus Q(n) is the length of the longest chain of nodes in Ent Set g.

(7*) List all nodes not marked as understood lying at a maximal arc distance Q(n) + 1 from whatever nodes are marked as understood on occasion n + r and obtain look ahead uncertainty/correct belief estimates; H_0^*, θ_0^*, for each one.

(8*) On occasion n + r recommend as the next aim node whichever node (subset of nodes) has a maximal (have maximal) value (or values) of θ_0^* in the estimated set. However, the student is allowed to choose any node as aim. The heuristic may be strengthened only to the extent that nodes at a greater arc distance than Q(n) + 1 are prohibited.

Together with the allocation and possible reallocation of strategic type (based on the manifest learning strategy and the competence evaluations performed at stage (8) or stage (5*) depending upon the current strategy type allocation, these instructions constitute an uncertainty regulation heuristic. Obviously, refinements are possible and may be desirable. For example, the gross estimates of $\Delta\bar{H}$, $\Delta\bar{\theta}$, might be derived by a combination of sequential sampling procedures, rather than recording empirical data from the paths actually transversed. But however crude, the heuristic, in its existing form, is effective (results are given in Section 11).

For all that, there is one fundamental point which may disturb the reader. A student (acting under the CET heuristic) is not allowed to access a node marked as understood.

The main justification for this edict is our theory; and it allows no margin for error. In fact, there seems to be no error; once that topic relations are understood, they appear to remain understood provided the strict conversation is anchored on the same domain. But the 'no access' rule also prohibits the direct detection of contrary instances; as a result of which it is open to the serious criticism that the theoretical dogma is 'self fulfilling'.

It is, of course, possible to allow the student access to an understood node, even though this entails some intricate organising routines. Suppose this permission is given and suppose, also, that a topic relation, with a node marked understood by the CET heuristic turns out not to be understood. Finally, assume that the node in question (say, node i) is accessed by the student when he has just understood some superordinate topic relation with node j.

Under these circumstances (which do not seem to occur in practice) it is clearly necessary to require that the student shall re-explain the topic relation R_i and that the process is iterated, so far as necessary, up to node j standing for the topic relation R_j.

The essential theoretical point is that any re-explanation of R_i, which is offered, consists in a derivation of topic relation R_i from topic relation R_j and not from the original prerequisites for addressing R_i. The point is that the theory may be fallible. But, if it is, then, as part of the theory, R_i must be reconstructible from R_j by decomposition of the explanation offered for R_j.

As noted, this requirement is practically vacuous in the context of a genuine conversational domain. It is relevant if the conversational domain is approximated, or if there is interfering experience between experiments.

9. More Intimate Types of Tutorial Conversation Heuristic

The tutorial conversation heuristic can be elaborated in several

ways. The learning process under an aim node can be adaptively controlled by systems of the type discussed in Chapter 2 provided that certain ranking conditions are satisfied by all members of the Ent Set of Aim. If the workset has only one member, a running index of θ (obtained by periodic problem posing; via BOSS) is analogous to ρ and η is a cue information index on the (single) tutorial data file that is accessed. If, as is usual, the working on set, has several (m_i) members indexed $j_i = 1, \ldots m_i$, then the adaptive mechanism must be multi-dimensional; that is (Chapter 2) m_i subcontrollers operate on variables η_{ji} (the cueing indices of m_i tutorial data files) to maintain each of m_i variables θ_{ji} (analogous to the ρ_{ji}) at their steady state value ξ_{ji} which is periodically adjusted by an overall controller. As an alternative scheme, the η_{ji} may be identified with priority ratings for attention to the possible subgoals.

For the subject matter of probability theory, rather little is gained by these expedients, but for more diffuse subject matter, local adaptive control is likely to be as effective as it is in the context of simple skills.

One other possibility is currently under investigation. It must be recognised that the entailment structure and the relational network from which it is derived, represent only a sample of what may be known. Most likely, knowledge is a continuum; though the topology of a complete epistemological structure contains gaps representing unmemorable relations. If this hypothesis is accepted, then the transfer characteristics (evidenced in Figs. 9 and 10 by the fact that the initial value of θ increases as learning proceeds), are not entirely determined by the links exhibited in the entailment structure. But hidden relations between topic relations (of which even the student is usually unaware) may be estimated by look ahead uncertainties determined for nodes other than the immediate aim candidates. Empirically, such data is obtainable; though not all other-than-candidate topic relations are deemed intelligible. In principle, at least, this data could be used in giving improved advice that recommends aim nodes to maximize transfer of training from topic to topic.

10. Simple Feedback Regulated Instruction

Given an entailment structure it is possible to write simple feedback regulated training programmes to suit students of various dispositions; for example, those who are prone to learn serially and those who are prone to be holist.

These training programmes embody teaching strategies, either serialist or holist in type, replacing the learning strategies that would be adopted by any successful student. Further, progress is monitored (as in conventional "branching programme") by tests and explanations. Though it is impracticable to achieve the detailed compromise between strategy and disposition for which the tutorial conversation heuristic is designed, it is at least possible to ensure that a student works under a strategy that belongs to an appropriate class. The serialist and holist teaching strategies employed, closely resemble the learning strategies of Plates 6 and 7. They are arbitrary to the extent that it is possible to choose several paths through the entailment structure which are logically sufficient to inculcate the subject matter and have the property (serialist case) of presenting one step at once or (holist case) maximising the size of the workset. In our experiments we used the paths most often selected by CET conversation students.

These teaching strategies could, of course, be instrumented in many different ways; for example, in a text or on tape-slide modules. To achieve uniform presentation amongst students, teaching strategies were, in fact, administered through CASTE for experimental purposes.

11. The Gross Effect of Learning Conditions upon Performance

Although the experiments on the learning of probability theory (discussed in Section 1.2) were primarily intended to check out the CET heuristic and to examine the uncertainty regulation CET (one tutorial conversation heuristic) groups of students were also required to learn under feedback regulated instruction. Within these subgroups some students were submitted to matched treatment (i.e. if their pretest result was holist to the holist strategy, if serialist to the serialist strategy); other students were deliberately mismatched (pretest holist, to serialist strategy and pretest serialist, to holist strategy).

As a general control group, some students were given an open licence to use the facility; the "Free Learning" group. The salient distinction between "Free Learners" and conversational learners is that the free learners did not have to satisfy an understanding criterion in order to progress though a few of them, already versed

TABLE 4.3
Data distinguishing holists and serialists of conversational students

Student No.		Mean H*	Mean θ*	Mean size work on set	Gandlemuller test score, Max. score=30
1[a]		0.19	0.13	1.0	10
3[a]		0	0	1.0	14
6[a]		0	0	1.2	6
7[a]		0	0.12	1.0	17
10[a]		0	0	1.3	11
	Mean	0.04	0.05	1.1	11.6
	S.D.	−0.07	0.05	0.01	3.7
2[b]		0.31	0.17	3.0	29
4[b]		1.30	0.31	3.4	30
5[b]		0.17	0.19	2.8	30
9[b]		0.16	0.14	2.6	28
9[b]		0.41	0.22	2.4	27
	Mean	0.47	0.21	2.85	29
	S.D.	0.42	0.06	0.02	1.2

a Serialists —Gandlemuller test: score $\leqslant$ 20.
a Holists —Gandlemuller test: score $\geqslant$ 25.

in "learning to learn" did seek an understanding condition on their own account. A relatively minor difference (which perhaps gave rise to some aimless and wasteful floundering around) is that the free learning students received no advice statements.

The main results are set out in Tables 3, 4 and 5. Of these Table 3 relates the initial strategic type test result, for conversational students only, to the student's typing by criteria attached to CET operation. Table 4 exhibits a gross comparison between the free learning, fixed teaching strategy and CET Tutorial Conversation Students. The post-test score is based on a personally administered test containing 46 main topics in which the topic relations must be explained to the examiner. The test score depends primarily upon an ability to explain though there is a residual component, no more than 15% of the entire score, that could be ascribed to factual retention only. Table 5 is a statistical summary of the principle differences observed in terms of performance, retention and method of learning.

Unequivocally, the tutorial conversation heuristic is the most effective teaching instrument. Matched students, whether of serial

TABLE 4.4

Treatment	Free learning	Fixed strategy	CET (Tutorial conversation)
Number of students in group	10	16 (8 matched & 8 mismatched)	10
Post-test			
Score mean	53.2%	48%	96%
S.D.	25.6	22	5.6
Tutorial interval exclusing procedures			
Mean (min)	244.3	221	132.5
S.D.	72.5	63.2	18.2
Mean ΔH	0.43	0.62	0.57
S.D.	0.2	0.32	0.3
Mean $\Delta \theta$	− 0.62	−0.72	− 0.46
S.D.	− 0.24	−0.3	− 0.03
Mean Gandlemuller test score out of 30	18	21	20.1
S.D.	10	9.2	10.6
Mean numerical ability test			
Score out of 20	12	11	11.5
S.D.	2.5	3.1	3.6

or holistic disposition, fare quite well (there is no significant difference between holists and serialists) but mismatched students (of either disposition) hardly learn at all. On average the free learning students fare badly, though not quite so badly as mismatched students. But there are some outstanding exceptions; students with their own "internal teacher" who "know how to learn"; these individuals reach the general tutorial-conversation level.

TABLE 4.5
Principal features revealed by statistical comparison of groups of students M_1 = number in Group 1, M_2 = number in Group 2.
The following abbreviations are used: ConvH = Conversational Holist, ConvS= Conversational Serialist, FixH = Fixed Strategy Holist, FixS = Fixed Strategy Serialist, FixMa = Fixed Matched Strategy, FixMis = Fixed Mismatched Strategy, FS= Fixed Strategy, Entire Group, FL = Free Learning, Entire Group, Conv'= Conversational Students.
All tests are non-parametric, either Mann Whitney U or Wilcoxon Matched Pairs, as appropriate to the difference examined.

Variable used	Comparison and sense of difference	M_1	M_2	P value	Significance level (per cent)
Test Score	Conv > FL	10	10	0.001 > P	0.1
	Conv > FS	10	16	0.001 > P	0.1
	Conv > FixMa	10	8	0.001 > P	0.1
	Conv > FL	8	10	0.017 > P	2
	FixMa > FixMis	8	8	0.001 > P	0.1
	FixS > FixH	8	8	0.36 > P	None
Tutorial interval	FL > Conv	10	10	0.001 > P	0.1
	FS > Conv	16	10	0.001 > P	0.1
	FixMa > Conv	8	10	0.001 > P	0.1
	FL > MixFa	10	8	0.23 > P	None
	FixMis > FixMa	8	8	0.323 > P	None
	FixS > FixH	8	8	0.1647 > P	None
Mean H*	ConvH > ConvS	5	5	0.004 > P	0.5
Mean θ*	ConvH > ConvS	5	5	0.004 > P	0.5
Mean θ*	FixMa > MixFis	8	8	0.001 > P	0.1
Transfer of training data.					
Comparison of halves of learning process.					
Mean ΔH	Conv 1st Conv 2nd	10	10	0.0057 > P	0.5
Mean $\Delta \theta$	Conv 1st Conv 2nd	10	10	0.0057 > P	0.5